Dynamic Architecture of Multi-Domain Carboxylases

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Moritz Hunkeler

aus Altishofen, LU

Basel, 2016

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Prof. Dr. Maier, Timm
Prof. Dr. Wahl, Markus C.

Basel, den 8.12.2015

Prof. Dr. Schibler, Jörg
Dekan

I Abstract

Biotin-dependent carboxylases are ubiquitous enzymes that play pivotal roles in key biosynthetic pathways. They employ a mobile biotin carboxyl carrier protein (BCCP) to catalyze ATP-dependent two-step carboxylation reactions utilizing two distinct enzyme activities, biotin carboxylase (BC) and carboxyl transferase (CT). Their substrate specificity is either towards small organic molecules, such as urea and pyruvate, or towards acyl-Coenzyme A (acyl-CoA) esters.

One of the most prominent members of acyl-CoA carboxylases, acetyl-CoA carboxylase (ACC), catalyzes the conversion of acetyl-CoA to malonyl-CoA, the highly regulated committed step of fatty acid biosynthesis. Eukaryotic ACC is a giant multienzyme, encompassing all catalytic domains, the BCCP and a large non-catalytic central domain (CD) on one type of polypeptide chain. The overall structure of eukaryotic ACCs as well as the structure of the CD, which is unique to eukaryotic ACCs, and its role in ACC regulation, remain unknown.

This thesis provides a comprehensive characterization of the dynamic structure of fungal ACC by combining crystal structure determination, small-angle X-ray scattering (SAXS) and electron microscopy (EM). The crystal structure of yeast CD, accompanied by low-resolution data on larger fragments up to intact fungal ACCs, reveals that the CD acts as a multi-hinged link between the BC and CT domains. Phosphorylation has an impact on the dynamic architecture resulting in a unique “mechanical” control mechanism.

A novel class of prokaryotic multi-domain acyl-CoA carboxylases (YCCs) was discovered recently. They share the same domain organization with eukaryotic ACCs, but lack the CD. A hybrid model of the *Deinococcus radiodurans* YCC (Dra YCC), together with quantitative analysis of BC domain mobility, provides novel insights into active site structure, domain interactions and dynamic assembly of these prokaryotic multienzymes. The implications of our structural studies of yeast ACC and Dra YCC for multienzyme engineering are discussed.

In addition, the crystal structure and NMR analyses of the membrane-bound, dimeric bacterial extracellular foldase PrsA reveal a bowl-like crevice as the key structural element for binding and folding of unfolded proteins.

II Table of Contents

III List of Figures

IV Abbreviations

ACC	Acetyl-CoA carboxylase
ACP	Acyl carrier protein
ADP	Adenosine diphosphate
ATP	Adenosin triphosphate
AMPK	5’ AMP-activated protein kinase
BC	Biotin carboxylase
BCCP	Biotin carboxyl carrier protein
BME	β-Mercaptoethanol
BT	BC-CT interaction domain
CD	Central domain
CDK	Cyclin-dependent kinase
ChREBP	Carbohydrate-responsive element-binding protein
CT	Carboxyl transferase
CoA	Coenzyme A
CPT-I	Carnitine palmitoyltransferase
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EM	Electron microscopy
GCC	Geranyl-CoA carboxylase
HEPES	4-(2-hydroxyethyl)-1-piperzineethansulfonic acid
hPar14	Human parvulin-like peptidyl prolyl *cis*/*trans* isomerase
FAD	Flavin adenine dinucleotide
FAS	Fatty acid synthase
kDa	Kilodalton
LCC	Long-chain acyl-CoA carboxylase
MALS	Multi angle light scattering
MCC	3-methylcrotonyl-CoA carboxylase
NAD	Nicotinamide adenine dinucleotide
NCBI	National center for biotechnology information
Ni-NTA	Nickel nitrilotriacetic acid
NMR	Nuclear magnetic resonance
NRPS	Non-ribosomal peptide synthase
o/n	Over night

PAGE	Polyacrylamide gel electrophoresis
PC	Pyruvate carboxylase
PCC	Propionyl-CoA carboxylase
PCP	Peptidyl carrier protein
PDC	Pyruvate dehydrogenase complex
PEG	Polyethylene glycol
PKA	Protein kinase A
PKS	Polyketide synthase
PPIase	peptidyl prolyl somerase
PT	Pyruvate carboxylase tetramerization domain
SAXS	Small angle x-ray scattering
SEC	Size exclusion
SDS	Sodium dodecyl sulfate
SREBP	Sterol regulatory element-binding protein
SurA	Survival protein A
TEV	Tobacco etch virus
TF	Trigger factor
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol
UA	Urea amidolyase
UC	Urea carboxylase
XL-MS	Cross-linking mass spectrometry
YCC	Acyl-CoA carboxylase

1 Introduction

1.1 Proteins, enzymes and metabolic pathways

Proteins account for one of the three cornerstones of the central dogma of molecular biology, which states that the genetic code is first transcribed into RNA and subsequently translated into proteins. Proteins are large macromolecules consisting of single or multiple long linear chains of amino acid residues. In all organisms, proteins are the main workers and responsible for diverse cellular functions such as catalyzing metabolic reactions, DNA replication and transcription, cellular trafficking, signal transduction and the maintenance of structural integrity. To achieve these tasks, they can, amongst other things, bind with varying affinity to other proteins, to ribonucleic acids and to small molecules, they can sense electric current and potential, and they can form cavities and pores[1].

Enzymes are proteins that act as molecular biocatalysts. They are required to accelerate chemical reactions in cells to rates fast enough to sustain life. This is achieved by reducing the activation energy of the reaction so it can take place at ambient temperatures. The ability to lower the activation energy is based on their three-dimensional fold that is encoded in the gene sequence itself. Enzymes are involved in almost all metabolic processes and in their entirety catalyze more than 5000 reaction types in all organisms[2]. They are specialized molecular machines that are highly specific to their cellular task and they can be regulated by diverse factors, leading to inhibition or activation[1].

The biosynthesis and degradation of complex metabolites cannot be achieved by catalysis of one single reaction but requires several sequential chemical reactions in a row, where the product of one reaction is the substrate for the next. This series of reactions is called a metabolic pathway and all the sequential reactions are catalyzed by enzymes, transforming the initial metabolite into the product. In compartmented eukaryotic cells metabolic pathways are located in all compartments and some pathways are distributed between compartments.

Metabolic pathways may be categorized into two classes[3]: First, those that produce intermediates that can also be used in other pathways, as for example glycolysis, the sequential breakdown of sugars into pyruvate. The intermediate

glucose-6-phosphate can also branch off and enter the pentose phosphate pathway or it can be converted to glycogen for storage. Second, pathways in which only the end products are of relevant biological function. A simple example for a pathway where the intermediates cannot be used is protein translation, since the only function of the intermediate is to become a protein. Many metabolic pathways are connected via their substrates and products (and as explained sometimes also their intermediates) and this flexible network can be tightly regulated and adapted to the environmental needs of a cell or organism. Through this tight regulation they ensure homeostasis within an organism. Metabolic pathways are normally mono-directional, one famous exception being again the glycolysis, where the direction of the pathway can be reversed and then be used to produce glucose from non-carbohydrate substrates such as pyruvate, glycerol or lactate.

In prokaryotes, the series of enzymes responsible for catalyzing all the steps in a pathway are often encoded in the same operon. The use of operons allows translation of polycistronic mRNAs which ensures a fixed stoichiometry of the respective enzymes[4]. Before and after every reaction in the active sites of the involved enzymes, the intermediate substrates need to diffuse from one protein to the next (Figure 1.1). However, free diffusion of reaction intermediates allows no control over efficiency and selectivity and the intermediates are in equilibrium with the bulk solvent. In addition, reaction intermediates can be highly reactive, unstable or insoluble, and may require specific support or shielding from the environment. In several metabolic pathways, small helper proteins are employed, which bind to the reaction intermediates, increase the efficiency and selectivity and provide protection[4].

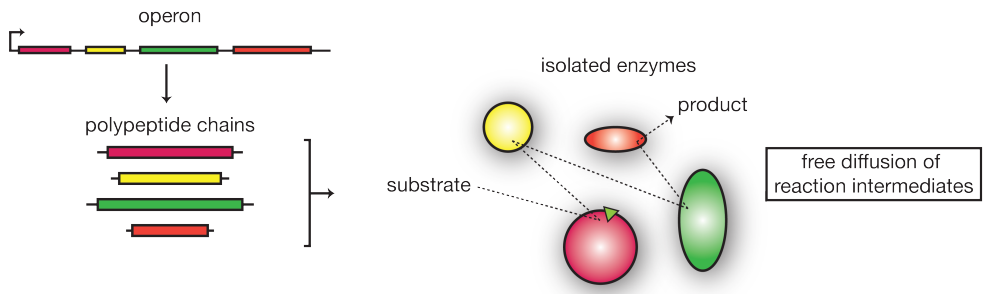


Figure 1.1 Proteins involved in metabolic pathways are organized in operons.
Schematic representation of the genome organization and the catalytic action of isolated enzymes involved in a multi-step reaction.

1.2 Carrier proteins shuttle and protect reaction intermediates

These small, so-called carrier proteins consist of only approximately 60 to 80 amino acids. Bound to carrier proteins, the selectivity and affinity of an intermediate towards the next active site in the pathway can be increased[5] and, according to the 'hot potato' hypothesis[6], unstable intermediates are sheltered on their reaction path and protected from aggregation. Binding to the carrier protein occurs via a prosthetic group that is attached to a conserved residue in the carrier protein (Figure 1.2A).

One prime example of a small carrier protein is the acyl carrier proteins (ACP, Figure 1.2B) involved in fatty acid biosynthesis, which consist of a four-helix bundle with a hydrophobic core. ACP is functionalized post-translationally with a 4-phosphopantheteine, derived from coenzyme A (CoA), at a conserved serine residue by the action of a dedicated phosphopantetheine transferase, the holo-acyl carrier protein synthase. During fatty acid biosynthesis, the growing fatty acid is covalently bound to the phosphopantetheine arm of ACP and, in *E. coli*, it was shown that the fatty acid chain is accommodated in the hydrophobic core as fatty acids are usually insoluble in an aqueous environment[7]. ACP shuttles the reaction intermediate between the active sites of the different enzymes working on it to produce a fully reduced fatty acid.

Peptidyl carrier proteins (PCPs, Figure 1.2B) are structurally similar to ACPs and are functionalized in the same way, by attachment of a phosphopantetheine moiety to a conserved serine. Analogous to ACP, PCPs are shuttling growing peptides in non-ribosomal peptide synthesis.

The lipoyl domain (Figure 1.2B), another carrier protein, consists of two small four-stranded β-sheets. It was originally discovered as a product of limited proteolysis of the E2-component of the pyruvate dehydrogenase complex[8,9] (PDC, see below). In the lipoyl domain, the N- and C-termini are located close together on one side and a highly conserved lysine resides in a loop on the other side of the protein. The lysine is post-translationally modified with its prosthetic group, lipoic acid, by a lipoate protein ligase. During pyruvate decarboxylation, it is responsible for the shuttling of an acyl group between different active sites.

Biotin carboxyl carrier protein (BCCP, Figure 1.2B) is structurally related to the lipoyl domain, consisting of a β-sandwich with quasi two-fold symmetry. As in the lipoyl domain, the N- and C-termini are on the opposing sites of the

conserved lysine that is covalently modified with biotin by a biotin ligase. BCCP is capable of shuttling a CO_2 group bound to its N1' nitrogen atom of the attached biotin and is utilized by carboxylases, decarboxylases and during transcarboxylation reactions[10], where it transfers the CO_2 moiety between two active sites.

Increasing selectivity of chemical reactions and sheltering of intermediates may in several cases, however, not be sufficient to ensure a fast and efficient reaction pathway.

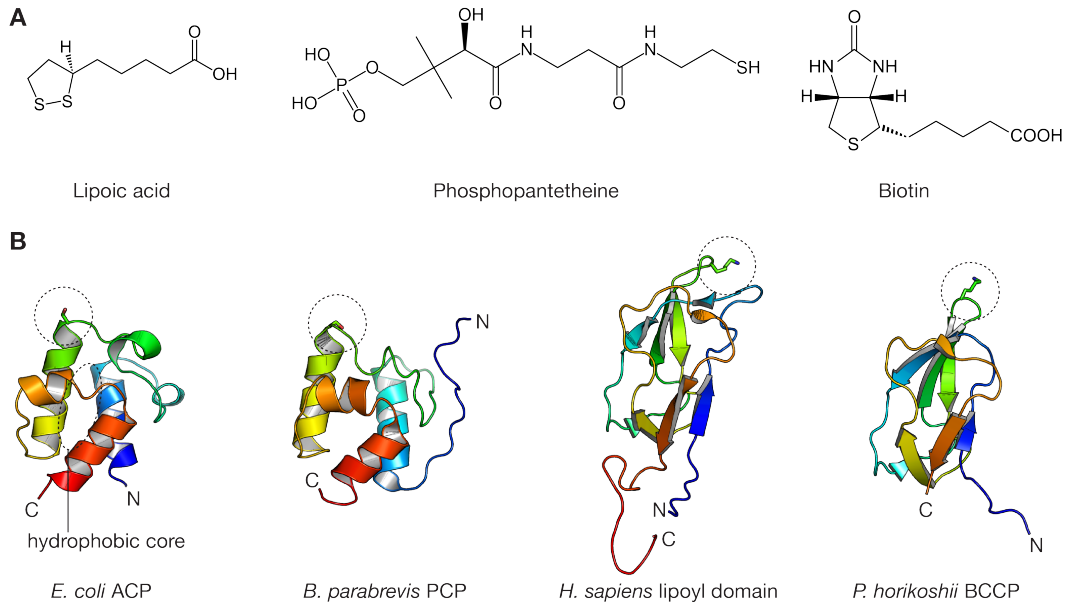


Figure 1.2 Overview of different cofactors and carrier proteins.
A. Chemical structure of the cofactors lipoic acid, phosphopantetheine and biotin. **B.** (from left to right) Crystal structure of FAS ACP from *Escherichia coli* (pdb: 2FAE). NMR solution structure of PCP (pdb: 2GDW) of the NRPS tyrocidine A synthetase from *Brevibacillus parabrevis*. NMR solution structure of Homo sapiens lipoyl domain from 2-oxoacid dehydrogenase. Crystal structure of *Pyrococcus horikoshii* BCCP (pdb: 2EVB) from a putative methylmalonyl-CoA decarboxylase. The functionalized residues are shown as sticks and marked with a circle and the hydrophobic core of ACP is labeled. Structural similarities between ACP and PCP and the lipoyl domain and BCCP, respectively, can be observed.

1.3 Integration of enzymatic functions into larger assemblies

The cell is a crowded environment[1], and as outlined before, some reaction intermediates need to be protected or shuttled and carrier proteins are one possibility to achieve this task. Additionally, relying only on diffusion may not always be efficient enough. Multi-subunit enzymes and multienzymes are

nature's way to increase the efficiency of catalyzed multi-step reactions where labile substrates are protected and either shuttled or channeled from one active site to the next.

1.3.1 Multi-subunit enzymes

Multi-subunit enzymes are ubiquitous and can be defined as 'non-covalent aggregates of protein subunits that catalyze successive steps of a multistep chemical reaction'[8]. Multiple enzymes are assembled into larger protein complexes held together by non-covalent interactions. Through the proximity of the individual proteins, only limited diffusion is required for the reaction intermediates in order to travel from one enzyme to the next, leading to increased speed and selectivity[11]. Metabolites can be 'channeled' between juxtaposed enzymes. Channeling was originally defined as the situation, when several sequential enzymatic reactions are performed on a substrate where the intermediates are out of the diffusion equilibrium compared to the same molecules in the solvent[3]. Direct interactions between sequential proteins also generate additional potential for allosteric regulation.

Depending on the lifetime of the protein-protein interaction, multi-subunit enzymes can be classified as transient or permanent[12] (Figure 1.3A). A *transient complex* associates and dissociates *in vivo* and is in a dynamic oligomeric equilibrium in solution with interactions forming and breaking continuously. The components of such transient complexes are initially not colocalized and thus require to be stable on their own, but the selectivity for the right partner has to be high. The equilibrium of dissociated and associated proteins can also be shifted by molecular triggers, such as a substrate binding. Transient interactions often occur in signal transduction pathways. For example, a heterotrimeric G-protein dissociates upon binding of guanosin triphosphate (GTP) into Gα and Gβγ, but is a stable heterotrimer when guanosin diphosphate (GDP) is bound. A further example for a transient multi-subunit enzyme, *E. coli* acetyl-CoA carboxylase (ACC), is discussed below. Ultimately, the interactions between subunits are driven by the localization and concentration of the individual subunits and the free energy of the complex, relative to the individual proteins. The interaction properties can thus be regulated by increasing or decreasing the local concentration of a subunit, by altering the localization or also by perturbing the

binding affinity. The affinity of a protein to another can be altered by binding of effectors like adenosine triphosphate or a change in physiological conditions.
In a *permanent complex* the protomers forming the complex cannot be observed as stable isolated structures *in vivo*. (Figure 1.3A) The protomers are often expressed simultaneously, for example under the control of the same promoter, facilitating colocalization. The protein interfaces are, in contrast to interaction areas in transient complexes, highly conserved. Examples for permanent complexes include the pyruvate dehydrogenase complex and various carboxylases and will be described in more detail in a later section.

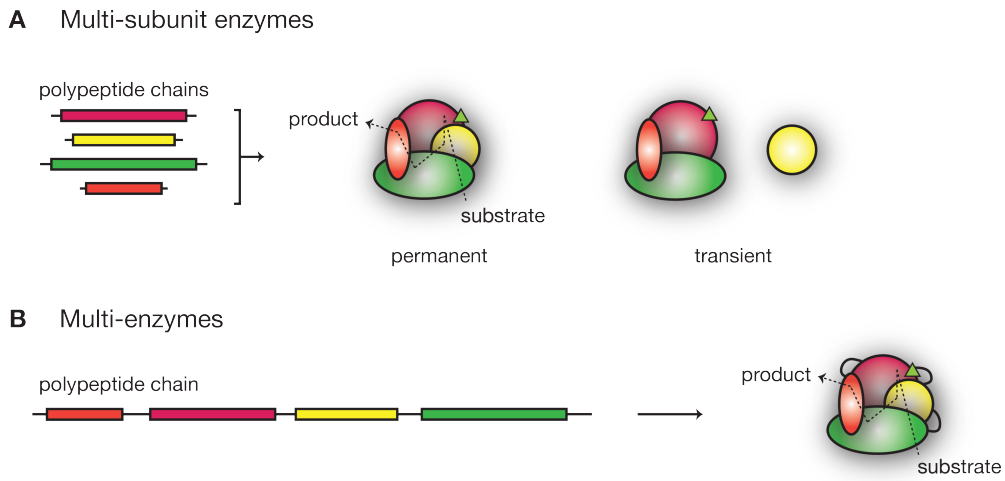


Figure 1.3 Difference between multi-subunit enzymes and multienzymes.
A. Schematic representation of the same reaction as in Figure 1, but catalyzed by a multi-subunit enzyme. In a permanent complex the individual enzymes are not stable in solution, in a transient complex one or all enzymes may dissociate from the complex. **B.** Schematic representation of the same reaction catalyzed by a multienzyme protein. In all cases, individual protein or subcomplexes can also oligomerize.

1.3.2 Multienzymes

In multienzymes, or multi-domain enzymes, several enzymatic components (domains) are encoded on a single polypeptide, as opposed to multi-subunit enzymes, where several catalytically active polypeptide chains assemble into one complex (Figure 1.3B). Multienzymes have often evolved by gene-fusion of separate multi-subunit ancestor proteins and occur more frequently in eukaryotes[13]. The single domains of such multienzymes show characteristics and properties of isolated enzymes and can sometimes be excised from the complex without loss of function. In other cases, domains are not active on their own without the interaction to their respective partner domains.

Multienzymes have the same advantages as multi-subunit enzymes, namely the potential to channel and protect metabolites, and the increased possibility for allosteric regulation. In addition, having several domains on one polypeptide helps to ensure a fixed stoichiometric ratio in eukaryotes and coordinated regulation of expression. It also raises the possibility to accompany the catalytic domains by additional non-catalytic domains or extended linker regions. These additional domains may be used for structural purposes like providing a scaffold for the correct positioning of domains or to provide reaction cavities that are shielded from the cytoplasm. Additional non-catalytic domains can also have regulatory purposes, by providing docking platforms for effector molecules and proteins or by harboring sites for post-translational modification.

Higher eukaryotic multienzymes often have simpler, independent analogues in prokaryotes, which can be exemplified by the enzyme CAD (a combination of the domain names: **C**PSase, **A**TCase and **D**ihydrooratase) that is involved in the *de novo* biosynthesis of pyrimidines. The first three enzymatic activities in this metabolic pathway are a glutamine-dependent carbamoyl phosphate synthase (**C**PSase), aspartate transcarbamoylase (**A**TCase) and dihydrooratase (**D**). In higher eukaryotes, these three enzyme entities are fused into one polypeptide chain with a molecular weight of 243 kDa while the three enzymes are independent in prokaryotes. In fungi, on the other hand, CPSase and ATCase enzymes are fused, but the dihydrooratase is encoded on a separate polypeptide[14].

Between the two mentioned extremes, multi-subunit enzymes and multienzymes, there are also intermediate assemblies, where some enzymatic domains are fused on a single polypeptide, that then interacts (in a stable or transient fashion) with additional proteins, for example the fungal fatty acid synthase multienzyme complex (discussed below).

In the following, multi-subunit enzyme and multienzyme refer to their strict definitions, as outlined above, while the terms ‘enzyme complex’ may be used for both cases. ‘Multi-domain enzyme’ will be used synonymously with ‘multienzyme’. The use of ‘multienzyme complex’ refers to an assembly of several distinct polypeptide chains where at least one of them carries more than one catalytic domain.

1.4 Carrier proteins in multienzymes

In order to ensure highly efficient shuttling of reactive or unstable intermediates, some multienzyme families also employ carrier proteins. These are essential for the proper function of a multienzyme by shuttling of the intermediate substrates without the risk of loss and protected from the influence of competing enzymes.
Before entering a more detailed description of different carrier protein-dependent multi-domain enzymes and multienzymes, some general consideration about flexibly tethered carrier proteins are outlined.

1.4.1 Tethering of carrier proteins provides mobility and restrictions

The prosthetic group of a carrier protein, together with the amino acid residue it is bound to, can be regarded as a swinging arm, and using simple geometry it can be calculated how far the 'business end' of the arm can travel when the domain itself is held in fix position. This consideration becomes important when the carrier protein is fused into a larger polypeptide and is tethered to this complex on at least one side. If the carrier protein is stably incorporated into the quaternary complex, the active sites that require to be served by the carrier protein have to be separated not farther than the distance the swinging arm can travel. Lipoic acid and biotin, bound to lysine, account for a swinging arm of about 16 Å, while the phosphopantetheine arm spans approximately 20 Å.
If, however, rotation of the domain around its anchor point(s) is accounted for, this maximal distance would increase approximately by the diameter of the domain. During the catalysis of multi-step reactions, this would lead to architectures where the carrier protein resides approximately in the middle of the complex, while the active sites are located around it[15]. In the examples provided below it becomes evident, that distances between active sites in multienzymes are often exceeding the maximum distance that could be spanned by the swinging arm of a carrier domain. This leads to the conclusion that the entire domain has to move in order to accomplish its task, which has been termed 'swinging domain model'[4].
The length of the polypeptide linkers connecting the carrier domains to the multienzymes is defining the maximum range of movement and by this restrains their diffusion so they only reach the active sites that they are targeting. The linkers, however, should not be considered as simple molecular strings [8], as they

can be rich in alanines, prolines and glycines. ^{13}C-NMR spectra of such linkers demonstrate that all of the Ala-Pro bonds (>95%) are in the all-*trans* conformation, indicating a stiffening of these regions as compared to random-coil structures, for which about 80-85% all-*trans* would be expected[16,17]. 95% all-*trans* correspond to an extended structure that allows substantial movement but also prevents entangling of the linker chains[18,19]. The length of these linkers is typically around 10-20 amino acid residues, which could span a maximal distance of approximately 30 to 60 Å in their most extended form, adding to the distance that the swinging arm can travel.

1.4.2 Lipoic acid-dependent enzymes

The lipoyl domain carrier protein is used in multienzymes responsible for the oxidative decarboxylation of 2-oxo acids, exemplified by the pyruvate dehydrogenase complex (PDC)[20,21]. PDC consists of 3 subunits (E1, E2 and E3) and catalyzes the conversion of pyruvate into acetyl-CoA, which then enters the citric acid cycle in mitochondria, and one to three lipoyl domains are tethered to E2. The first enzyme in this complex (E1) is a thiamine diphosphate-dependent decarboxylase that catalyzes the decarboxylation of pyruvate and acylates the lipoyl group. E2 also harbors a dihydrolipoyl acetyltransferase that catalyzes the transfer of the acyl group from the lipoyl domain to CoA, yielding acetyl-CoA. The reduced lipoyl group is then regenerated at the active site of E3, a dihydrolipoyl dehydrogenase. During this process the lipoyl domain shuttles between the active sites (Figure 1.4A).

E2 is the core protein and assembles either in octahedral (24-mer, Gram-negative bacteria) or in icosahedral (60-mer, Gram-positive bacteria and eukaryotes) symmetry. In addition, it contains a small binding domain responsible for binding to E3, and in some organisms also to E1[22]. Up to twelve E3 subunits and up to 24 copies of E1 can bind tightly but non-covalent, via linker domains, to the octahedral core[22]. To the icosahedral core up to 60 E1 or E3 subunits can bind. The bound E1 and E3 subunits can adopt a variety of conformations and span a maximum range of motion of approximately 150 Å[20]. In eukaryotes, an additional protein, E3BP, binds to the core, and further proteins, such as specific kinases and phosphatases involved in regulation, may also bind. The whole complex can reach a molecular weight of up to 10 MDa, including a large number of carrier proteins shuttling between the core and the peripheral enzymes.

1.4.3 Phosphopantetheine-dependent enzymes

Phosphopantetheine-dependent enzymes include fatty acid synthase (FAS), polyketide synthase (PKS) and non-ribosomal peptide synthase (NRPS). 4'-phosphopantheteine is bound covalently to a conserved serine residue in the acyl carrier protein[23] (ACP, in the case of FAS and PKS) or the homologous peptide carrier protein[24] (PCP, in NRPS). FAS is responsible for the biosynthesis of long-chain fatty acids[1]. In *E. coli*, this is orchestrated by a series of individual enzymes, but in many other organisms the enzymes have been incorporated into multienzymes. FAS works iteratively[19], meaning that the domains are reused in a circular fashion, increasing the size of the fully reduced growing fatty acid by two carbons in each round. It uses acetyl-CoA as starting molecule and malonyl-CoA for the subsequent elongation steps. The growing fatty acid is bound to ACP, which shuttles it throughout this whole process from one domain to the next, until it is released at the very end of the synthetic process.

Fungal FAS is a barrel-shaped 2.6 MDa $\alpha_6\beta_6$ hetero-dodecamer with two large reaction chambers, where the different enzymatic components are divided onto two polypeptide chains. ACP is tethered flexibly on two sides in the center of the formed cavity. Due to the hetero-dodecameric assembly, there are three ACPs in the upper reaction chamber and three in the lower. The linkers, N-terminally stiffened through a high content in alanine and proline to reduce entanglement, limit the free diffusion of ACP to the closest set of active sites, that are all facing the inside of the barrel. The distance covered between the linker attachment point and the sulfhydryl group of the phosphopantetheine arm is approximately 55 Å, while the distance between active sites range from 60 to 80 Å[25]. These distances could be covered by little translation and a rotation of ACP using the linker attachment sites as a pivot point. However, it is currently unclear how exactly substrate shuttling is achieved.

Mammalian FAS is an X-shaped homodimeric protein with all the seven proteins fused into a single polypeptide chain including the carrier protein ACP (Figure 1.4B). ACP can be considered to be tethered on one side only, since the linker to the following highly mobile C-terminal thioesterase domain is extensive (25 amino acids). In the X-shaped homodimer, all active sites are lining the reaction clefts with inter-domain distances of up to 85 Å, too large to be explained by a swinging-arm model of the phosphopantetheine. Therefore, large domain movements are required for proper function of the multienzyme[25,26]. In addition to

the movement of ACP, FAS is highly flexible, with the upper and the lower part only connected by two short polypeptide linkers. The reaction clefts can open and close and the upper part of the molecule can rotate and shift relative to the lower part[26,27]. In contrast to ACPs from prokaryotes, animal ACP does not bury the growing acyl chain in its core[28].

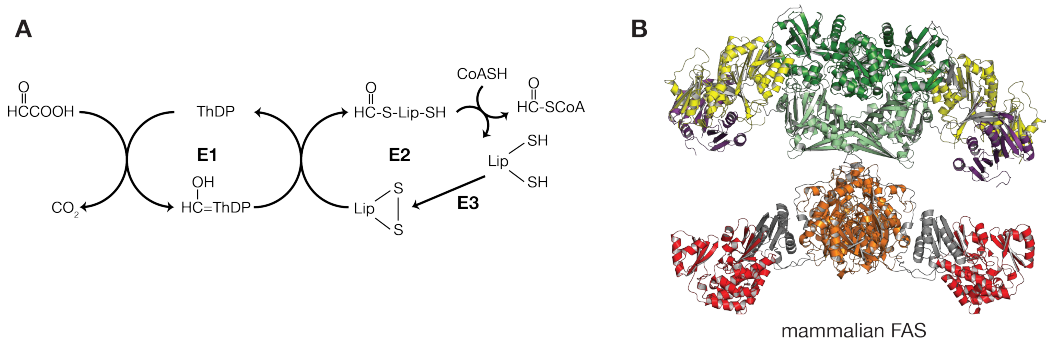


Figure 1.4 PDC and FAS, enzyme complexes involved in metabolic pathways.
A. Simplified schematic depiction of the reaction catalyzed by PDC. The transfer of an acyl group by the lipoyl domain is shown. The enzymes are labeled as described in the main text. ThDP: thiamine pyrophosphate, Lip: lipoyl domain, CoA: Coenzyme A. **B.** Crystal structure of the mammalian FAS (pdb: 2VZ8) as a paradigm of multienzyme organization. Different domains of this dimeric protein are colored in different colors.

PKS and FAS are related enzymes, consisting of the same core domains and using similar building blocks. While FAS exclusively synthesize fully reduced fatty acids, PKS have a wide variety of products, producing secondary metabolites, such as anti-cancer agents, immunosuppressants and antibiotics[29-31]. Like FAS in *E. coli*, all enzymes can be encoded by single genes and work as isolated proteins, or they can be joined in multi-functional polypeptides[8]. Iterative PKS work similar to FAS, but they may be arranged in a different order or specific domains may be doubled or left out. Modular PKS, on the other hand, utilize every domain only once, and in order to catalyze a complex multi-step reaction, many enzymes are required. These enzymes are organized in so-called modules, where every module consists of several domains and is responsible for a certain modification step in the biosynthesis pathway. Every module has an independent ACP, which is determined to shuttle the growing polyketide within its own module and to hand it over to the next module. The individual modules are organized in a beads-on-a-string fashion, resembling an assembly line, and there is clear colinearity between the genetic context of the domains and the synthesis of the product, e.g. the first domains that are translated are also the first to

engage a substrate. The product structure is thus directly encoded in the modular arrangement of the coding gene.

NRPS are responsible for a the biosynthesis of a diverse set of peptide secondary metabolites, such as toxins, siderophores, pigments, antibiotics or immunosuppressants[32]. This large variety of non-ribosomal peptides (NRP) is achieved by the consecutive condensation of amino acids in an mRNA-independent manner in a process that is not limited to the 20 proteinogenic amino acids. NRPS are large modular multienzymes, and every single condensation step is catalyzed by a distinct set of domains, again organized in modules. Following the reactions scheme established for FAS and PKS, the growing substrate is bound to the phosphopantetheine arm of PCP and shuttled from one module to the next, where it is handed over to the next carrier protein. As for PKS, depending on the domain composition of the module and the elongation substrate, different modifications take place. NRPS mostly occur in bacteria and fungi, and they can be multi-subunit complexes translated from operons or multienzyme complexes[33].

1.4.4 Biotin-dependent enzymes

Biotin is used as cofactor in three different types of enzymatic reactions: First, in the ATP-dependent carboxylation of small organic substrates by biotin-dependent carboxylases[34]. Second, in the decarboxylation of β-keto acids and their thioesters by biotin-dependent decarboxylases and, third, in the interconversion of oxaloacetate and propionyl-CoA into pyruvate and methylmalonyl-CoA, respectively, by transcarboxylase[10]. Biotin-dependent carboxylases are described in detail in later sections of this thesis.

1.5 Multienzymes as tools for synthetic biology and combinatorial biosynthesis

Polyketides offer a large biological and industrial potential, especially in the advent of the antibiotic crisis, with more and more bacteria developing resistance to available antibiotics. Blockbuster drugs like the antibiotics Zithromax and Biaxin and the immunosuppressants Rapamycin and Prograf sale for billions of dollars. These compounds are all complex polyketides, produced by modular PKS. Most naturally produced polyketides, however, are not working as drugs,

since they did not evolve to be active in human bodies. Additionally, the discovery of new compounds is a difficult and laborious task, which involves cultivation of soil-living bacteria and the chances of discovering a new antibiotic class are very small[35]. There is a need to not only discover new compounds, but also to modify them in order to, for example, increase their bioavailabilty[36]. Nature manages, however, to produce and modify compounds in large numbers, starting from simple precursors, utilizing the modular nature of PKS and NRPS.

In an approach called 'combinatorial biosynthesis', modular multienzymes, such as PKS and NRPS, are ideal templates for synthetic assembly lines, as they can catalyze the complicated multi-step synthesis of important biological compounds that are difficult to produce chemically in a lab[37]. Since the structure of the product is encoded in the genetic context, swapping of single domains on the gene level allows researchers to alter the outcome of the reaction and thus the nature of the produced compound[38,39]. Applying this technique on 6-deoxyerythronolide B synthase (DEBS), a PKS responsible for the synthesis of the antibiotic erythromycin, researchers were able to produce over 50 new unnatural compounds[40]. Engineering of PKS by point mutations and random mutagenesis has also been reported[41].

In order to construct new synthetic pathways, a large basis of structural and biochemical knowledge of natural multienzymes is desirable. Often single domains and parts of multienzymes are well characterized. However, the overall architecture of full-length enzymes including the position of flexible linkers, arrangement of docking and linking domains, domain interactions, substrate shuttling and regulation mechanisms lack proper characterization. Without these information and without the possibility to rely on structural data, the task of designing new synthetic enzymes becomes very difficult.

Biotin-dependent carboxylases are another example for carrier protein-dependent enzyme complexes, where individual parts are well studied but information on overall dynamic assembly modes, at least for the most prominent members, is sparse.

1.6 Biotin-dependent carboxylases

Biotin-dependent carboxylases were discovered more than 50 years ago and their biochemical properties and biological roles have been studied extensively.

They are widely distributed in all domains of life and they play pivotal roles in the metabolism of fatty acids, amino acids and carbohydrates[42-45]. In some microorganisms they are implicated with CO_2 fixation[46,47], methanol assimilation[48], acetyl-CoA assimilation[49-51], 3-hydroxypropionate assimilation[52], polyketide biosynthesis[53], utilization of urea as nitrogen source[54,55], metabolism of terpenoids[56] and mycolic acid and methyl-branched fatty acid biosynthesis[55].

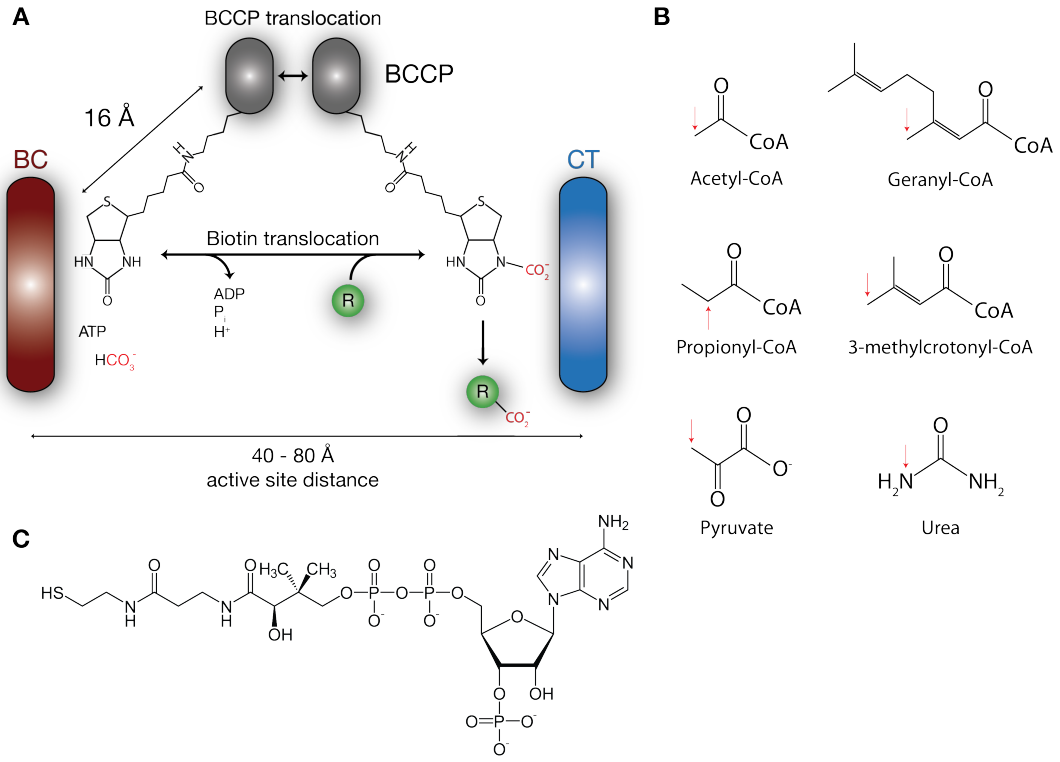


Figure 1.5 Mechanism and substrates of biotin-dependent carboxylases.
A. Schematic reaction scheme employed by biotin-dependent carboxylases. Biotin is covalently bound to BCCP (gray), which, after carboxylation in the active site of BC (ruby), shuttles carboxybiotin to the active site of CT (blue) for carboxyl transfer onto the substrate (R). Biotin and BCCP translocations are indicated by arrows and important distances are marked. **B.** Different substrates of biotin-dependent carboxylases. The site of carboxylation is marked by a red arrow. **C.** Chemical structure of CoA.

The reaction catalyzed by biotin-dependent carboxylases occurs in two half-reactions at two different enzymatic domains (Figure 1.5A): The biotin carboxylase (BC) component utilizes bicarbonate as CO_2-donor for ATP-dependent carboxylation of the N1' atom of a biotin cofactor that is covalently bound to a conserved lysine in the biotin carboxyl carrier protein (BCCP). Subsequently, the carboxybiotin is shuttled to the carboxyl transferase (CT) component where the CO_2 is transferred to an acceptor molecule. The acceptor (or substrate) is often a CoA-ester of organic acids and the site of carboxylation

is either the α-carbon or the γ-carbon. In addition, small compounds, such as pyruvate or urea, can serve directly as substrates[34]. During the reaction, biotin must visit the active sites of both BC and CT. Recent structures of full-length carboxylases demonstrate that the distance between active sites is in the range between 55 and 85 Å[57-61]. Based on a 'swinging arm' hypothesis[4] with limited rotation of BCCP, biotin could translate only by approximately 30 Å, showcasing that also carboxylases are dependent on a 'swinging domain' model.

Depending on enzyme and organism, the BC, BCCP and CT can be on separate polypeptide chains or partially or fully fused into one (Figure 1.6). While sequence conservation in the different carboxylases is high for the BC and BCCP, the CT component is less conserved and the sequence and active site architecture differ depending on the nature of the substrate.

Based on their substrate specificity, biotin-dependent carboxylases are classified in three groups: First, pyruvate carboxylase (PC) and second, urea carboxylases (UC) both act on small organic compounds while the third family of acyl-CoA carboxylases is specific for CoA esters of a variety of carboxylic acids (Figure 1.5B,C).

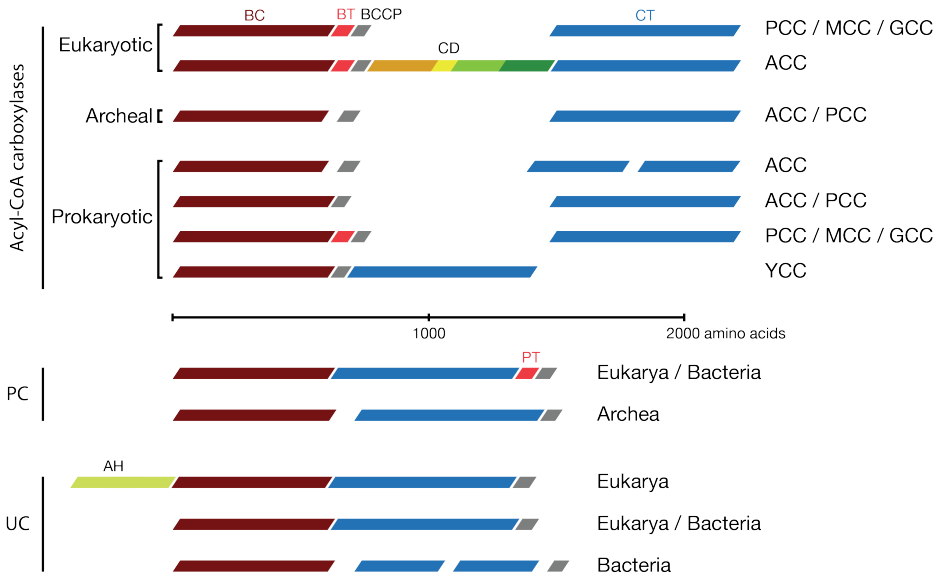


Figure 1.6 Domain organization of biotin-dependent carboxylases.
Overview of all families of biotin-dependent carboxylases, separated by class and kingdom of life. All individual domains are labeled and color-coded, the scale bar denotes the number of amino acids. All abbreviations are according to the main text.

1.7 Urea carboxylase and pyruvate carboxylase

In order to utilize urea as energy source, many bacteria, fungi and algae possess the ability to convert urea to ammonium[62,63]. In these organisms, UC catalyzes the first step in this pathway, namely the carboxylation of urea to produce allophanate, which is then hydrolyzed to generate ammonia and CO_2 by allophanate hydrolase (AH)[54,64]. In a subgroup of organisms, the AH can be fused to the UC, forming a multienzyme called urea amidolyase (UA, Figure 1.6). In the yeast *S. cerevisiae*, UA is part of the pyrimidine degradation pathway while in other fungi, such as *C. albicans*, it can be involved in the virulence. Here the released CO_2 is a signal for hyphal switching and the ammonia product is secreted to alkalinize the extracellular environment[65]. On a sequence level, UC can be a multi-subunit enzyme or a multienzyme (Figure 1.6).

A crystal structure of a 137 kDa UC component of an UA from the fungus *K. lactis* has been reported (Figure 1.7). It shows an elongated monomeric structure with the BCCP tightly bound to the CT domain. The active sites face each other and are separated by about 60 Å, consistent with a swinging domain model. The CT domain does not share sequence homology with other carboxylases, and consistently, its structure and oligomerization state is distinct. The quaternary structure of CT domains of other carboxylases is at least that of a dimer, which leads to very distinct protein architectures in the following examples of different families.

PC was discovered in 1960 in chicken and beef liver[66] and by now has been found in most organisms. It catalyzes the carboxylation of pyruvate yielding oxaloacetate[67,68] and in mammals it is involved in gluconeogenesis in liver and kidney, and glyceroneogenesis and lipogenesis in adipocytes (Figure 1.8). PC deficiencies are linked to lactic acidemia, hypoglycemia, psychomotor retardation and death[69,70]. In eukaryotes and most bacteria, PC is a multi-domain enzyme with a monomeric weight of approximately 130 kDa (Figure 1.6), with the BCCP tethered C-terminally to CT by a 14 amino residue long linker, functioning only as a tetramer. In archae and some bacteria, it can occur on two polypeptide chains, but the overall assembly mode is the same as in the multienzyme form. While the BC domain shows considerable conservation to other carboxylases, the CT component is again distinct in sequence.

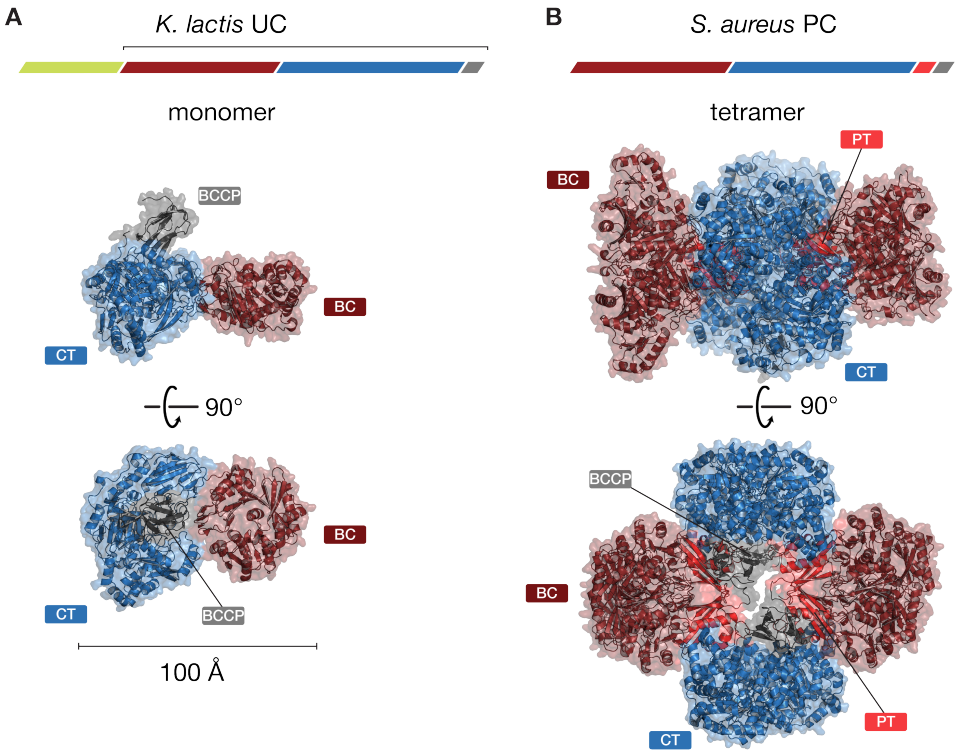


Figure 1.7 Architectures of UC and PC.
Crystal structures of monomeric UC from *Kluyveromyces lactis* (**A**, pdb: 3VAF) and the tetrameric PC from *Staphylococcus aureus* (**B**, pdb: 3BG5). The structure of the AH domain of UC was not determined. Color-coding according to the sequence scheme. Same domains are labeled where applicable and the molecules are to scale.

Crystal structures are available and show the PC tetramer in a square shape, with two protomers stacked in the 'bottom' layer and two in the 'upper' layer, roughly obeying 222 point group symmetry[61] (Figure 1.7). An additional domain was identified, that is important for the tetramerization of the complex, and thus was called PT (PC tetramerization). PT, BC and CT are all forming homodimers with domains in the other layer. BCCPs are located near the middle of the assembly, bound to the active site of a CT domain, inserting the biotinylated β-hairpin deep into a binding pocket. An interesting architectural feature is the fact that the BCCP-bound biotin is carboxylated by the BC domain of the own protomer and is then shuttled to the CT domain of the other protomer in the same layer, meaning that the carrier protein is not only serving its own protomer during catalysis. The respective active sites are separated by 75 Å, again indicating that a swinging domain mechanism of the BCCP has to be employed.

Strikingly different architectures have been presented for these two classes of biotin-dependent carboxylases, and this structural diversity will increase during the detailed description of acyl-CoA carboxylases in the next section.

1.8 The family of acyl-CoA carboxylases

The family of acyl-CoA carboxylases includes acetyl-CoA carboxylase (ACC), propionyl-CoA carboxylase (PCC), 3-methylcrotonyl-CoA carboxylase (MCC) and geranyl-CoA carboxylase (GCC), and the individual family members will be discussed below. Generally, all of these carboxylases are present in prokaryotes and eukaryotes, except GCC, which is only found in *Pseudomonas* and *Acetinobacter* strains[71,72] (Figure 1.6).

1.8.1 Overview of biological role of acyl-CoA carboxylases

PCC, located in mitochondria in mammals, converts propionyl-CoA to *D*-methylmalonyl-CoA and, in most organisms, it is important for the catabolism of β-branched amino acids (Thr, Val, Ile) and fatty acids with an odd number of carbon atoms by feeding them ultimately into the tricarboxylic acid cycle (TCA cycle, Figure 1.8). In several archaeal organisms and bacteria, it is also implicated in CO_2 fixation[46,47], methanol assimilation[48], acetyl-CoA assimilation[49-51], 3-hydroxypropionate assimilation[52] and mycolic acid and methyl-branched fatty acid biosynthesis[55]. Deficiencies in PCC activity lead to propionic acidimia, which may lead to death in severe cases[73]. In addition, PCC deficiency leads to epilepsy and seizure, implicating that PCC expressed in the brain may be involved in neurodevelopment[74,75].

MCC shares considerable sequence conservation with PCC, even though it is acting on γ-carbons while PCC acts on α-carbons. In eukaryotes, MCC is located in mitochondria and catalyzes the conversion of 3-methylcrotonyl-CoA to 3-methylglutaconyl-CoA and is essential for the catabolism of leucine and isovalerate in most organisms[76] (Figure 1.8). 3-Methylglutaconyl-CoA is ultimately turned over to acetyl-CoA and enters the TCA cycle. Deficiencies in MCC activity are linked to 3-methylcrotonylglycinuria, which manifests in a range from asymptotic individuals to severe cases of psychomotor retardation, seizure, coma and death[77,78].

GCC can be classified into the same class as MCC, since the amino acid sequences between the two are highly conserved and since they both work on a γ-carbon of a α-β unsaturated acid. In *Pseudomonas* and *Acetinobacter* strains, GCC accepts geranyl-CoA and converts it to γ-carboxygeranyl-CoA and is involved in the metabolism of the geranyl group and other acyclic terpenes[56].

Owing to this activity, some bacteria can use geranyl as only carbon source in a process similar to β-oxidation, where they break down γ-carboxygeranyl-CoA to 3-methylcrotonyl-CoA (also releasing two acetyl-CoAs and acetate), which is then a substrate for MCC and is further degraded.

ACC catalyzes the carboxylation of acetyl-CoA, yielding malonyl-CoA, the dedicated substrate for fatty acid biosynthesis and a potent inhibitor of β-oxidation. The detailed biological role of ACC is discussed below.

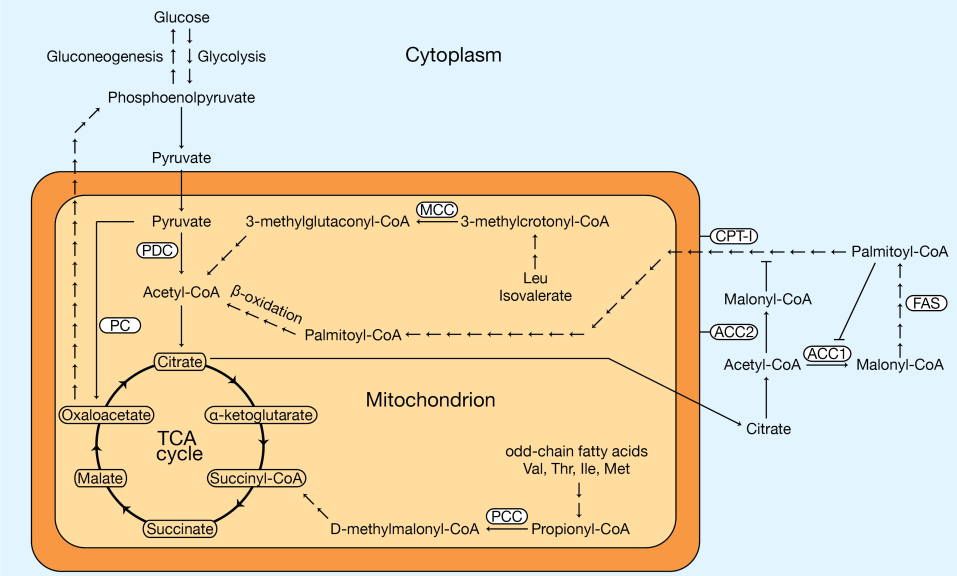


Figure 1.8 Functions of biotin-dependent carboxylases in mammalian cells.
Important pathways that involve biotin-dependent carboxylases in mammalian cells are marked with the name of the corresponding enzyme. Multiple arrows indicate pathways where more than one enzyme is involved. Selected intermediates of the tricarboxylic acid cycle (TCA cycle) are shown. Malonyl-CoA, produced by ACC1, is the committed substrate for fatty acid biosynthesis while malonyl-CoA produced by ACC2 acts as inhibitor of CPT-I. All abbreviations according to the main text.

1.8.2 Architectural features of acyl-CoA carboxylases

All acyl-CoA carboxylases share the principle composition of BC, BCCP and CT, with well conserved BC and BCCP components and a more variable CT. In some cases, additional non-catalytic components are added to this core set (Figure 1.6). A BC-CT interaction domain (BT) can be found in all eukaryotic members of the family as well as in prokaryotic PCC, MCC and GCC[58,59]. Eukaryotic ACC contains a large additional central domain (CD) that is biochemically and structurally poorly characterized, even though phosphorylation of certain residues in the CD has been implicated in regulation of ACC[79-81].

Even though the domain composition is shared amongst all acyl-CoA carboxylases, the relative domain organization and stoichiometry differs vastly between the members of the family. Prokaryotic acyl-CoA carboxylases are multi-subunit enzymes, which consist of two to four subunits. Eukaryotic acyl-CoA carboxylases can either be multi-subunit proteins, as for example the PCC, or multienzymes, such as the ACC, combining all domains on a single polypeptide chain of about 2300 amino acids[34] (Figure 1.6).

Structures of the intact assemblies of these enzymes are only available for the bacterial MCC[58] and PCC[59] (Figure 1.9A,B). Both proteins share the same domain arrangement on a sequence level and show the same stoichiometry, forming $\alpha_6\beta_6$ hetero-dodecamers with a size of approximately 750 kDa. The β-subunits (CT) form a hexameric core by trimerization of CT-dimers, similar to the 12S subunit of transcarboxylase[82], with three α-subunits (combined BC, BT and BCCP) attached at the top and bottom of the CT ring. While both enzymes obey 32-point group symmetry, a striking rearrangement of the α subunits relative to the β subunits can be observed. Overall, PCC forms a compact, cube-shaped assembly, while the MCC is more elongated. Additionally, the hexameric organization of the subdomains in the β subunits and the structure and placement of the BT domains differ. BC domains in PCC are monomeric with contacts only to the BT and the CT, while in the case of MCC they associate into a trimer and show no contact to the CT ring. The active sites of BC and CT are separated by 55 Å in PCC and 80 Å in MCC, repeatedly underlining the fact that the carrier protein has to undergo large-scale conformational change to position the biotin cofactor in both active sites.

Even though the large architectural difference between PCC and MCC was initially unanticipated, it fits well with the observation of large organizational freedom even within the family members of acyl-CoA carboxylases. This is exemplified in the case of ACC. Bacterial ACC consists of a transient assembly of four subunits following a $(BC)_2(BCCP)_4(CT\alpha CT\beta)_2$ stoichiometry[42], while biochemical studies suggest that archaeal ACC works in a $(BC)_4(BCCP)_4(CT)_4$ stoichiometry[83]. Eukaryotic ACC, on the other hand, is a large multienzyme with BC, BT, BCCP, CD and CT combined on one polypeptide chain with a monomer weight of approximately 250 kDa and functioning either as dimers or as higher oligomers[84,85].

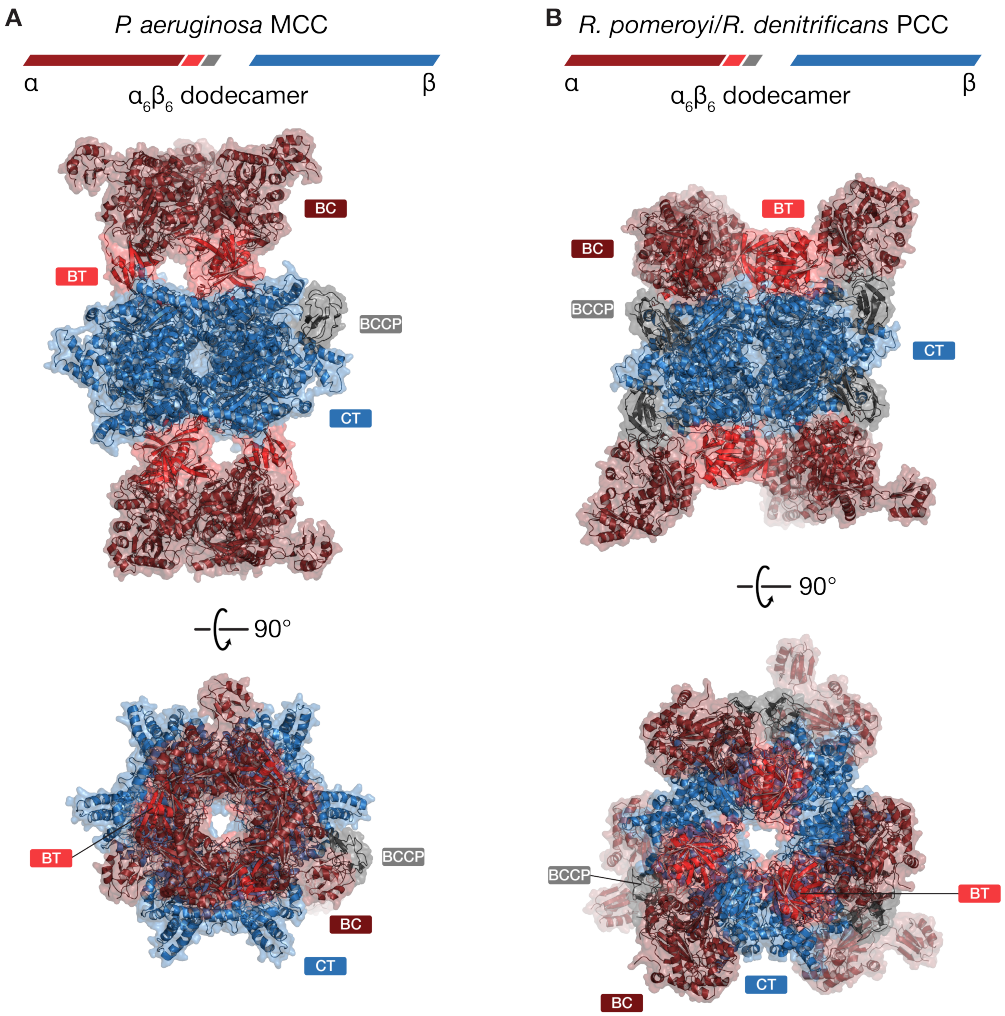


Figure 1.9 Architectures of MCC and PCC.
Crystal structures of hetero-dodecameric MCC from *Pseudomonas aeruginosa* (A, pdb: 3U9S) and hetero-dodecameric chimeric PCC from *Ruegeria pomeroyi* and *Roseobacter denitrificans* (B, pdb: 3N6R). Color-coding according to the sequence scheme. Same domains are labeled and the molecules are to scale.

While structures of bacterial biotin-dependent carboxylases[58-60,86] and of individual domains from eukaryotic ACCs are known[79,87,88], there is no insight into the architecture or assembly mode of full-length eukaryotic ACC and this cannot be deduced from the already available structures. Additionally, there is only little understood on the function and positioning of the CD.

1.9 Biological role, regulation and domain structures of acetyl-CoA carboxylase

1.9.1 Biological role of ACC

ACC was discovered in 1958 due to its involvement in fatty acid biosynthesis[89]. There are two isoforms of ACC in mammals, ACC1 and ACC2. While ACC1 is mostly expressed in lipogenic tissues, such as liver, adipose and lactating mammary gland, ACC2 is predominantly expressed in heart and skeletal muscle. ACC1 catalyzes the committed and rate-limiting step of long-chain fatty acid biosynthesis[44,90-92], as the produced malonyl-CoA is mainly used as two-carbon donor by FAS (Figure 1.8). ACC2 shares 73% sequence identity to ACC1, but contains an N-terminal anchoring sequence that locates it to the outer mitochondrial membrane. Malonyl-CoA produced by ACC2 is a potent regulator of fatty acid oxidation in these tissues by inhibiting the carnitine palmitoyltransferase I (CPT-I, Figure 1.8)[93]. CPT-I is responsible for the conversion of long chain fatty acids into acylcarnitine, a conversion that is essential for the fatty acids in order to cross the mitochondrial membrane and to enter β-oxidation[94-96]. Mice deficient in ACC2 have continuous β-oxidation, reduced level of body fat and are protected against diabetes and obesity induced by high-fat diets[97,98].

Single nucleotide polymorphism in human ACC is linked to hypertriglyceridemia and hypercholesterolemia in patients taking antipsychotics[99]. Also, human ACC is overexpressed in cancer cells and chemical inhibition or RNAi knockdown of ACC have been shown to lead to increased levels of apoptosis and growth inhibition in cancerous cells[100-104].

Fungi contain only one cytosolic ACC isoform and it is essential (Figure 1.8). Mutants lacking ACC arrest in G2/M phase[105], while mutants lacking FAS can survive if long-chain fatty acids are provided in the medium[106]. Adding long-chain fatty acids does not restore viability of ACC^- mutants[107] and this indicates that ACC may also be involved in the synthesis of very long-chain fatty acids. Yeast mitochondria also encode an isoform of ACC, called HFA1, crucial for fatty acid and lipoic acid biosynthesis in mitochondria[108-110].

In some bacteria, such as *Streptomyces coelicolor*, the produced malonyl-CoA is additionally used for polyketide biosynthesis[111], and in archae, ACC is thought to be involved in CO_2 fixation[83].

1.9.2 Regulation of acetyl-CoA carboxylase

ACCs are tightly regulated on the level of expression, by binding of proteins and small molecules and by post-translational modification. The features and differences of regulation between human ACC and the simpler model organism *S. cerevisiae* will be discussed in this section.

Long-term regulation is achieved by changing the number of active enzymes in a cell or compartment. Several transcription factors regulate gene expression of human ACC, including sterol regulatory element binding proteins (SREBP1a and SREBP1c) and carbohydrate response element binding protein (ChREBP). High-carbohydrate diets lead to elevated levels of insulin, which in turn upregulates expression of SREBP1c and ChREBP, while high levels of polyunsaturated fatty acids decrease the expression of SREBP1[112-114]. In addition, the effects of these transcriptional factors can be further increased by PGC-1α and PGC-1β, peroxisome proliferator-activated γ coactivator 1[115,116].

The use of different promoters leads to alternative splicing at the 5' of human ACC1. The shortened N-termini lack phosphorylation sites that are normally recognized by AMP-activated protein kinase (AMPK)[117-120]. Phosphorylation in these regions is linked to activity, and it is proposed that the shortened isozymes are constitutively active. Similarly, ACC2 also has a shortened transcript, and the encoded protein is believed to lack the N-terminal anchoring sequence, rendering the enzyme cytosolic [117].

In yeast, steady state transcription is decreased by the phospholipid precursors inositol and choline[105,121,122]. Adding inositol and choline to the growth medium leads to a 50-75% reduced activity level of ACC, based on decreased levels of mRNA[123].

Short-term regulation of human ACC is achieved by feedforward and feedback regulation (Figure 1.10A). Citrate, a precursor of acetyl-CoA, serves as allosteric activator, while long-chain acyl-CoA, the product of fatty acid biosynthesis, acts as an inhibitor[45,93,124]. The most active forms of animal ACCs are large, linear polymers with a molecular weight of up to 8 MDa, consisting of ACC dimers as building blocks[45,125] (Figure 1.10C). Citrate and long-chain acyl-CoA are either driving or perturbing this polymerization process, respectively (Figure 1.10A). Binding of the small protein MIG12 to ACC is further promoting polymerization of both mammalian ACC isoforms[84]. However, an N-terminally truncated construct for ACC2, lacking the mitochondrial membrane anchoring sequence, was used in

this study, and it was not shown whether ACC2 also polymerizes when it is membrane-bound.

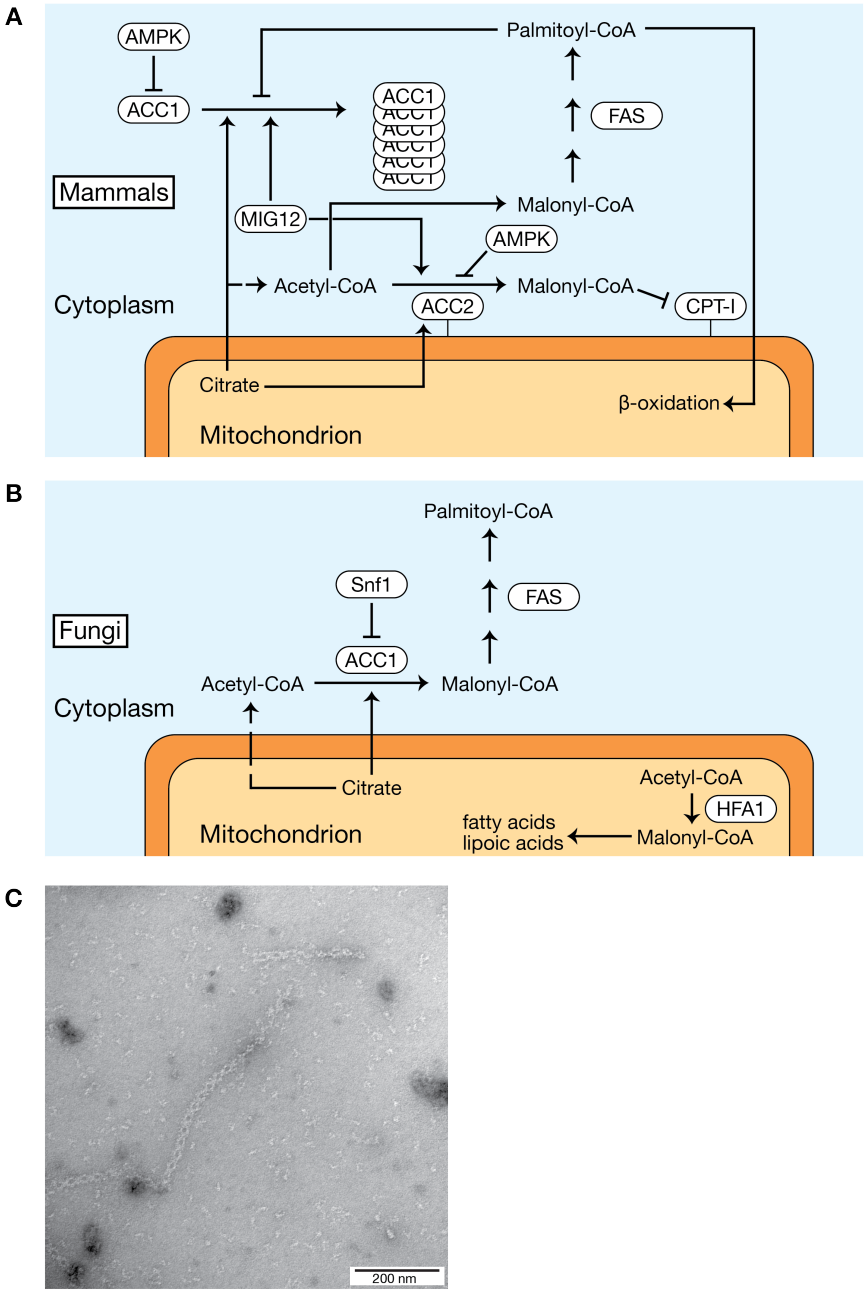


Figure 1.10 Regulation of ACC in mammals and fungi.
Feed-forward and feedback regulation of ACC is indicated by arrows. **A.** ACC1 and ACC2 are activated by citrate and MIG12, which leads to polymerization in ACC1. Phosphorylation by AMPK leads to inactivation of ACC, as does binding of palmitoyl-CoA. **B.** The fungal AMPK homologue Snf1 inhibits ACC by phosphorylation, while the acetyl-CoA precursor citrate is an activator of activity. Malonyl-CoA produced by mitochondrial HFA1 is used for the biosynthesis of mitochondrial fatty acids and lipoic acids. **C.** Negative stain EM micrograph of *Homo sapiens* ACC1 polymers (unpublished results). Scale bar is indicated.

While long-chain acyl-CoA has no inhibitory effect on the activity of yeast ACC, citrate has also been identified to induce activity[45]. However, no spontaneous or inducible polymerization has been observed for yeast ACC (Figure 1.10B).

Human ACC is also regulated by reversible phosphorylation (Figure 1.10A) by AMPK and cAMP-dependent protein kinase (PKA). AMPK phosphorylates ACC1 *in vitro* at Ser80, Ser1201, and Ser1216 and PKA at Ser78 and Ser1201. Direct regulatory effects, however, have only been observed for phosphorylation of Ser80 and Ser1201[126,127]. Ser222 in ACC2 aligns to Ser80 in ACC1, and is also phosphorylated by AMPK. Ser80 is highly conserved in higher eukaryotes and presumably binds, upon phosphorylation, into the Soraphen A (an antifungal and antibiotic compound) binding pocket[128]. Ser1201 is only moderately conserved among higher eukaryotes, while Ser1216 is highly conserved among all eukaryotes. However, no effect of Ser1216-phosphorylation on activity has been reported.

Breast cancer susceptibility gene 1 (BRCA1) binds to phosphorylated Ser1263 and this binding decreases ACC1 activity and establishes a further direct link between ACC and cancer[129,130]. Phosphorylation of Ser1263 is controlled by the cell cycle, possibly through cyclin-dependent kinase (CDK).

The situation presents itself less complex in yeast (Figure 1.10). Several phosphosites were identified in phosphoproteomics studies, namely Ser2, Ser735, Ser1148, Ser1157, and Ser1162. Of these phosphorylation sites, only Ser1157 is highly conserved among all fungi and it aligns to Ser1216 in human ACC. Phosphorylation of Ser1157 by the AMPK homologue Snf1 leads to strongly reduced activity *in vitro* and *in vivo*[80,131,132].

Many of the phosphorylation sites and also the binding site for BRCA are located in the central domain of yeast and human ACC, respectively. This domain, however, lacks structural description and the mechanisms underlying regulation through phosphorylation are not understood, neither in humans nor in fungi.

1.9.3 Structures of eukaryotic acetyl-CoA carboxylase

ACC holo-enzymes are difficult to study. In prokaryotes, the multi-domain enzyme is a transient complex that readily dissociates during purification while eukaryotic ACCs are large polypeptides and a flexible quaternary structure has to be expected, at least for the mobile carrier protein. Thus, no structural data is available for the complete assembly of eukaryotic ACCs and the structure of the

CD is unknown and has no homologs in structural databases. Crystal structures of the individual domains, however, have been solved.

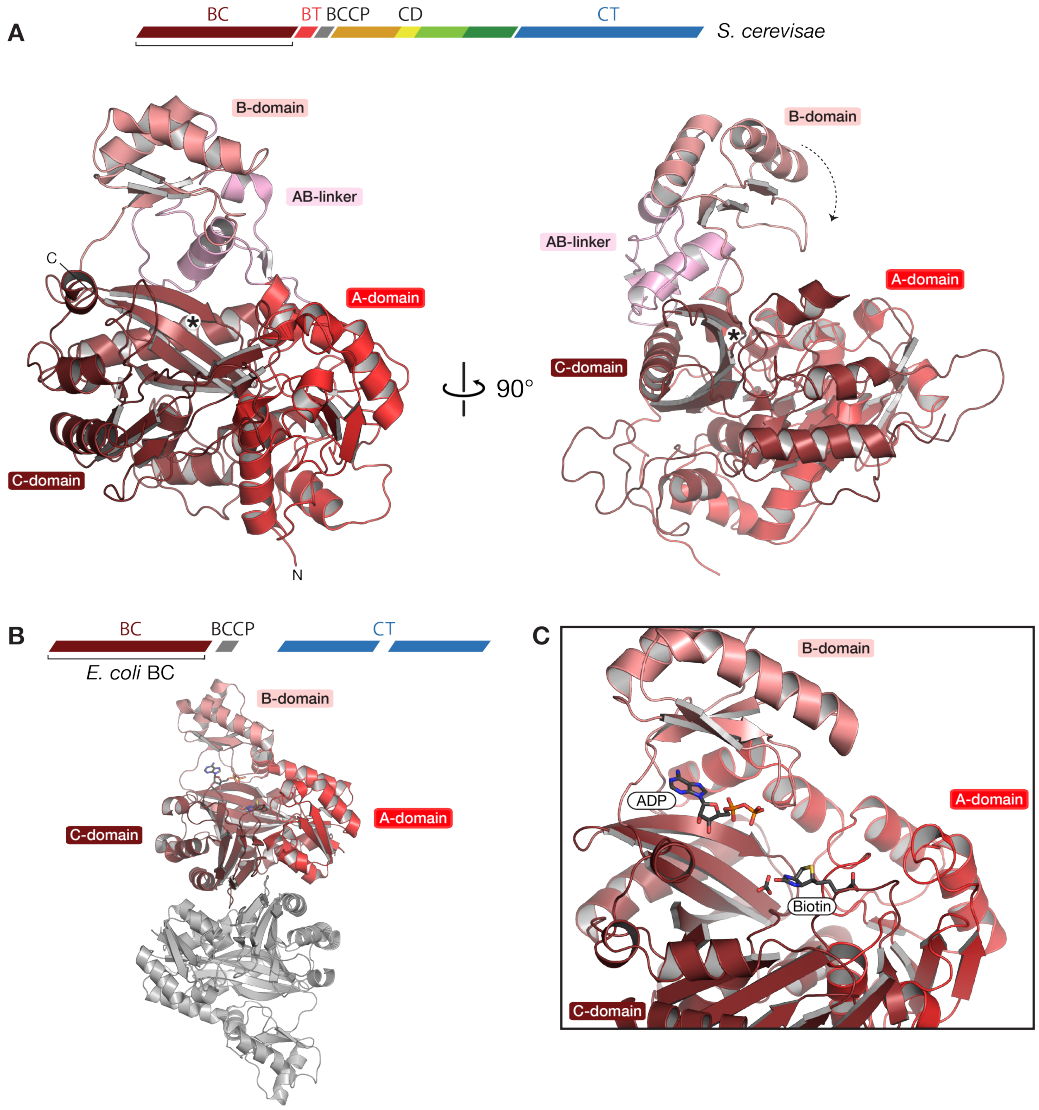


Figure 1.11 Structure of the biotin carboxylase domain.
A. Crystal structure of the monomeric *S. cerevisiae* BC domain (pdb: 1W93). The A-, B-, C- and AB-linker domains are colored and labeled with different shades of red. The active site is marked with an asterisk, and the closing movement of the B-domain is indicated with a dashed arrow. **B.** Crystal structure of the dimeric BC domain from *E. coli* (pdb: 3G8C) in complex with ADP, biotin and carbonate (shown as sticks). The subdomains labeled as in A. **C.** Close up on the active site of *E. coli* BC, illustrating the positioning of ADP, biotin and carbonate (as sticks). The subdomains are labeled as in B.

The BC domain of eukaryotic ACC is monomeric in solution while prokaryotic BC is dimeric (Figure 1.11A,B). It is suggested, that conformational differences in the respective dimerization area (compared to *E. coli*) prevents dimerization [133]. However, eukaryotic BC is inactive in solution, and it remains concealed whether BC requires the proximity of other domains in the complex for activity or a

whether specific conformational change has to occur. Apart from that, the overall structure clearly resembles the BC subunit of *E. coli* ACC, consisting of three domains (A, B and C). Eukaryotic BC contains several inserted segments, especially the linker segment between the A- and the B-domain (AB-linker). Residues from the A- and C-domain contribute to the active site and the flexible B-domain potentially acts as a lid that closes over the active site during catalysis [134]. The positioning of the substrates can be derived from a comparison with *E. coli* BC and is shown in Figure 1.11B,C.

During catalysis, a nucleophilic attack by one of the oxygen atoms of carbonate on ATP leads to formation of a carboxyphosphate intermediate, which decomposes and releases orthophosphate. Orthophosphate acts as general base and extracts a proton from the N1' atom of biotin, and the resulting enolate biotin is then stabilized by a positively charged arginine during the carboxylation reaction.

A crystal structure of yeast CT was reported in 2003[88] (Figure 1.12A). As in their prokaryotic counterparts, the CT domain is a head-to-tail dimer consisting of two subdomains (N-lobe and C-lobe) of which each shows a crotonase fold. In the structure, CoA is bound and it can be seen that residues from both protomers are required for binding and recognition (Figure 1.12B). The pantetheine arm wedges into a cavity between the two protomers and places the thiol group in a pocket formed by both protomers. Since the active site is situated exactly between the protomers, CT requires dimerization in order to be active and is consequently not stable as a monomer in solution. Interestingly, it is proposed that the C-terminal helical regions are incompatible with a hexameric assembly, as it was observed for CT domains of bacterial acyl-CoA carboxylases [135]. This strongly suggests that eukaryotic ACC show a completely unrelated assembly to the bacterial counterparts.

The CD is a completely unique feature of eukaryotic ACCs and is predicted to consist of an α-helical part, which accounts for roughly half of the CD, followed by one or more domains with mixed α-helix and β-sheet predictions. It has no homologs in other proteins, no other carboxylase incorporates a similar domain and no biochemical or structural data is available, except for the fact that it is implicated in regulation through phosphorylation.

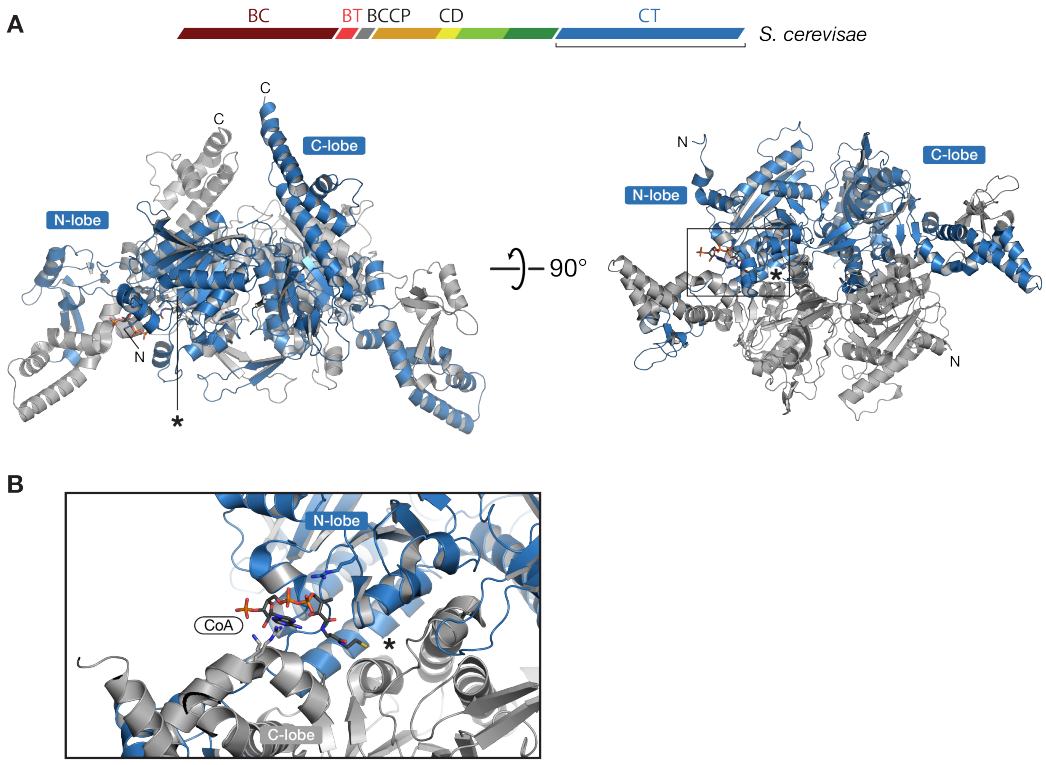


Figure 1.12 Structure of the carboxyl transferase.
A. Crystal structure of the dimeric *S. cerevisiae* CT domain (pdb: 1OD2). The N- and C-lobes are labeled and one of the active sites is marked with an asterisk. Bound CoA is shown as sticks. Coloring of one protomer according to the sequence scheme, the protomer is in gray for clarity. Visible N- and C-termini are labeled. **B.** Zoom-in on the active site (marked area in A), situated between the N-lobe of the first protomer and the C-lobe of the second protomer. The bound CoA and the coordinating residues, from both protomers, are shown as sticks. The expected site of catalysis is marked with an asterisk.
ACC as target for drug development

1.10 ACC as target for drug development

Due to their pivotal role in fatty acid metabolism, ACCs are a potent target for drug development against type 2 diabetes, atherosclerosis, cancer, and microbial and fungal infections[90,124]. Many of these diseases are manifestations of the metabolic syndrome, which in turn is linked to obesity[136]. It has been projected that by the year 2030 approximately 50% of the adult population in the US will be obese[137].

The potency of ACC as a drug target was first demonstrated by the observation that ACC2 knock-out mice exhibit elevated β-oxidation, reduced body fat and weight and improved insulin sensitivity[97,98]. A complete knock-out of ACC1, on the other hand, leads to embryonic lethality in mice[138], while a liver-specific knock-out of ACC1 in mice results in reduced lipid accumulation without disturbing glucose homeostasis[139]. Recent reports, however, failed to observe

these effects[140-142], and it is suggested that differences in the strategies that were used to produce the knock-out mice are responsible for this discrepancy[34].

Many potent inhibitors of human ACC are known[100,143-148]. Some of them do not discriminate between the two isoforms, while others show increased selectivity towards ACC2. In rodent models, these inhibitors show similar effects as it was seen in the knock-out mice studies: increased β-oxidation, decreased body fat mass and improved insulin sensitivity. Although long-term treatment using isoform non-selective inhibitors leads to increased β-oxidation, it does not sustainably reduce hepatic triglyceride levels in rodent models[149]. Prolonged inhibition of ACC1 reduces glucose-stimulated insulin secretion. ACC2-selective inhibitors may thus be more suited as potential drugs against diabetes[150].

ACC is overexpressed in different cancers, such as liver, breast and prostate cancer, rendering the enzyme an attractive target for the development of new anti-cancer compounds. RNAi knockdowns and chemical inhibition lead to growth arrest, apoptosis and decreased tumor cell invasion[100,103,104,151,152].

Two commercially available herbicides, in use for over 20 years, are inhibitors of plastid ACC: aryloxyphenoxypropionate and cyclohexanedione[44]. Not all plants, however, are sensitive to these herbicides and the widespread use lead to the emergence of new resistant weeds. The two compounds are only weakly active against mammalian and yeast ACCs[153]. The macrocyclic antifungal and antibiotic compound Soraphen A is a potent inhibitor of eukaryotic BC, but does not target the prokaryotic counterpart[87,133].

Inhibiting one of the two half-reactions occurring in ACC is sufficient to inhibit the overall reaction and most available inhibitors target the CT, while only soraphen A is targeting the BC. Inhibition of polymerization could be an additional route for potential drugs in higher eukaryotes, while selective targeting of the CD could potentially work on all eukaryotes. For this task, however, more knowledge is needed concerning the structure and function of the CD.

1.11 Bacterial single-chain acyl-CoA carboxylases

A novel group of bacterial acyl-CoA carboxylases was recently discovered but not characterized biochemically[34]. Even though they are denoted as PC or carbamoyl-phosphate synthetase in the NCBI database, sequence homology to acyl-CoA carboxylases clearly identifies them as members of the latter. These

carboxylases, denoted as YCCs here, consist of a single polypeptide chain comprising all three functional domains in the same order as eukaryotic ACCs but they, however, lack the large CD (Fig 2D). The monomer weight is approximately 120 kDa but no extensive studies concerning the oligomerization state or the architecture have been conducted. The domain arrangement on a single polypeptide is exceptional as they are the only true bacterial acyl-CoA multienzyme carboxylases. YCCs are present in a diverse set of unrelated bacteria, such as *Pseudomonas aeruginosa*, *Cupriavidus metallidurans*, *Deinococcus metallidurans* and *Mycobacterium avium* subspecies *paratuberculosis*. For most YCCs neither substrate specificity, which could differ vastly, nor cellular function have been studied. Recently, a crystal structure and initial substrate characterization has been reported for an YCC from *Mycobacterium avium*, which was termed long-chain acyl-CoA carboxylase (MapLCC, Figure 1.13). Again, this structure shows a new overall architecture: The CT domains form a trimer of dimers, as it has been seen in other acyl-CoA carboxylases (Figure 1.9A,B), but the dimeric BCs are attached laterally. Further structural studies of these enzymes could shed light on the dynamic architecture, modes of substrate shuttling and domain interactions with significance for all multi-domain carboxylases.

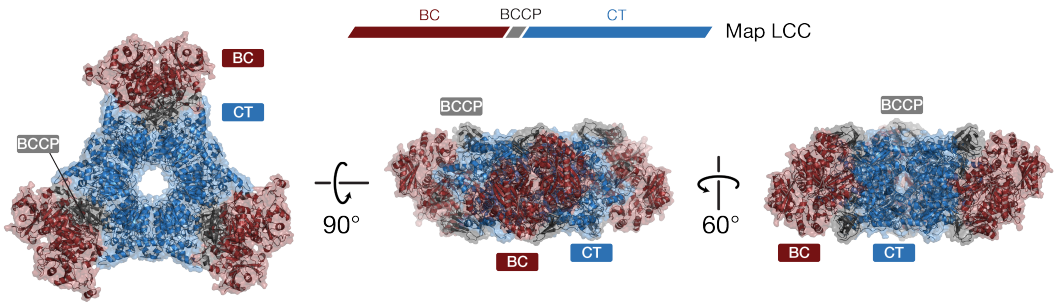


Figure 1.13 Crystal structure of MapLCC.
MapLCC consists of a hexameric CT ring with dimeric BC domains laterally attached. Views correspond to top view and side views focusing on the BC dimer or the CT dimer. Domains form one polypeptide chain are labeled and colored according to the sequence scheme (pdb: 4RCN).

1.12 Aims of the thesis

As outlined so far, there is profound knowledge on the biochemistry of multi-domain acyl-CoA carboxylases available, especially ACC, gathered over the last 60 years. ACCs are of outstanding importance in primary and intermediary metabolism and they are established as important drug targets. However, structural information is limited to individual domains and insights into the molecular mechanisms of regulation, substrate shuttling, domain interactions and dynamic architecture of these multienzymes are sparse. Structure determination was precluded due to the size and flexibility of these fascinating enzyme complexes and without structural details it remains elusive how exactly the enzymes operate on a molecular level.

Therefore, the main aim of this thesis was to analyze the structure and function of eukaryotic acetyl-CoA carboxylases, in particular the currently uncharacterized central domain. I wanted to understand the biological role of the CD, especially how it can mediate regulation through phosphorylation and what implications this may have for other carrier protein-dependent enzymes. In addition, I strived to gain knowledge on the architecture and function of the full assembly and domain dynamics, in order to understand how multienzymes work.

At the beginning of the project, *D. radiodurans* YCC was a completely uncharacterized enzyme that was misannotated in all databases. The work on YCC was conducted in order to characterize this particular prokaryotic multi-domain acyl-CoA carboxylase functionally and structurally and to provide insights into active site architecture, domain interactions and dynamics that may hold true not only for this enzyme but also for a larger class of carrier protein-dependent multienzymes.

To achieve this, a variety of methods were used, based on x-ray crystallography as a main tool, supplemented with electron microscopy (EM), small angle x-ray scattering (SAXS) and additional methods for biophysical characterization.

Chapter 2 of this thesis covers our findings on eukaryotic ACC and is based on a manuscript in revision. Several crystal structures of full-length and truncated eukaryotic ACCs will be presented and a link will be made on how this enzyme can be regulated by phosphorylation of a single site in the central domain. This novel mechanical regulation mechanism gives rise to new possibilities in synthetic biology.

Chapter 3 outlines what we learned from structural analyses of the prokaryotic YCC from *D. radiodurans* concerning large conformational flexibility in multi-domain carboxylases. This part is based on a manuscript in preparation. A hybrid structure and quantitative analysis of domain flexibility reveal a highly flexible architecture and demonstrate that biotin-dependent multienzymes can exhibit large conformational freedom.

In addition, I was involved in crystal optimization, data collection and structure refinement in an unrelated side-project, which will be introduced below. The findings of this work are presented in chapter 4, which is reproduced from manuscript publication in J. Biol. Chem.[154]. It will provide insights into the structure and function of the bacterial extracellular foldase PrsA using a combination of crystallographic, enzymatic and NMR spectroscopic analyses.

1.13 The bacterial extracellular foldase PrsA

The rate-limiting step in the folding of proteins containing *cis*-prolines is catalyzed by peptidyl prolyl *cis*/*trans* isomerases (PPIases). PPIases are foldases that occur in all types of cells[155]. These proteins are ubiquitous and can be divided into three classes: Cyclophilins, FK506-binding proteins (FKBP) and parvulins[156]. Cyclophilins and FKBPs are targets for immunosuppressants and are thus also called immunophilins. The name-giving prototype for the parvulin family is an enzyme from *E. coli*[157]. This small cytoplasmic protein consists essentially of only the PPIase domain. In other cases, a homologous PPIase domain can be flanked by additional protein stretches either N- or C-terminally or on both termini[158]. Based on functional and structural differences, at least three subfamilies can be defined, that are discussed below.

The first class of the Pin1-type contains, except in plants, an N-terminal WW-domain[159] (Figure 1.14A). WW domains are known to bind to proline-rich protein stretches and to phosphoserine- or phosphothreonine-containing motifs[160]. A crystal structure of the human Pin1 protein reveals that the PPIase domain consists of a four-stranded antiparallel β-sheet surrounded by four α-helices ($\beta\alpha_3\beta\alpha\beta_2$)[159]. The WW domain is shown to indeed bind together with the PPIase domain to phosphorylated residues located N-terminally to a proline[161]. Pin1-type PPIases are involved in mitosis and chromatin remodeling[162,163].

In the second subclass, hPar14-type, the first α-helix and the preceding loop (s1-H1 loop) are shorter than in Pin1-type PPIases and the exact biological function is unknown[164] (Figure 1.14A,B).

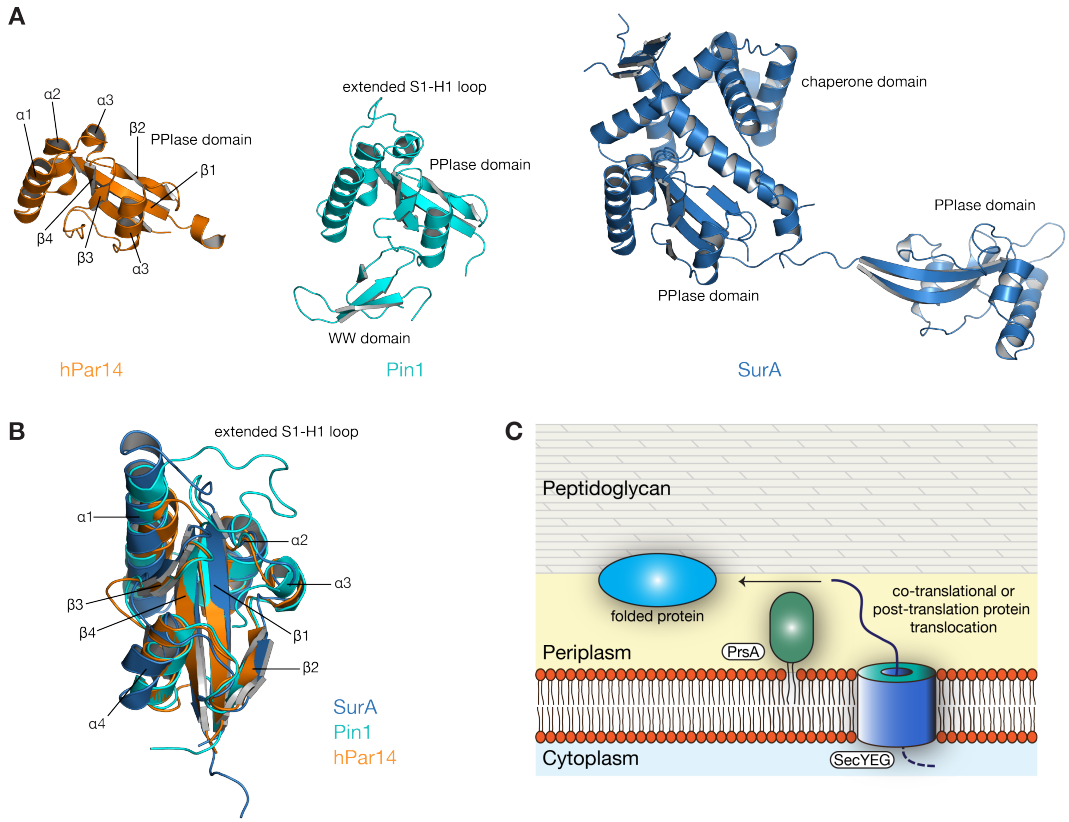


Figure 1.14 PrsA is an extracellular foldase belonging to the parvulin family.
A. Crystal structures of representatives of the three classes of parvulins: hPar14 (orange, pdb: 3UI4), human Pin1 (cyan, pdb: 2Q5A) and *E. coli* SurA (blue, pdb: 1M5Y). All structures are to scale and aligned on the PPIase domain. The secondary structure elements are labeled for hPar14, as well as all domains. Pin1 contains an additional WW domain and SurA contains two PPIase domains and an additional chaperone domain. **B.** Structural overlay of the PPIase domains from hPar14 and Pin1 and the core PPIase domain from SurA, illustrating the common fold and the extended S1-H1 loop of Pin1. **C.** Schematic representation of extracellular folding, mediated by the membrane-anchored foldase PrsA. SecYEG is the main protein-conducting channel in gram-positive bacteria.

In the third subclass, the PPIase domain is most similar to the *E. coli* parvulin and it is occurring mostly in prokaryotes, required for maturation and folding of exctracytoplasmic proteins[165]. The periplasmic chaperone SurA, for example, belongs to this class (Figure 1.14A).

Parvulins, in general, assist the maturation of intracellular proteins, and a large network of periplasmic PPIases support the folding of periplasmic and outer membrane proteins in Gram-negative bacteria[166,167]. In Gram-positive bacteria,

folding and maturation of most secreted proteins is dependent on a single parvulin-type PPIase, PrsA[168,169].

PrsA is an essential protein in *Bacillus subtilis* under normal growth conditions[170] and depletion of PrsA from cells leads to increased osmotic shock susceptibility[171], decreased cell wall integrity[172,173] and increased antibiotic sensitivity[174,175]. PrsA is anchored to the outer leaflet of the cell membrane at the interface between plasma membrane and cell wall[168,169]. It is crucial for proper folding of proteins that are protruding from the protein-conducting channel SecYEG during co-translational or post-translational protein translocation (Figure 1.14C). PrsA is a potential target for drug development due to its important role as folding factor of secreted proteins[168], toxins[176] and virulence factors[174,177]. On a sequence level, 155 N-terminal and 70 C-terminal amino acid residues flank a parvulin-like domain. These flanking regions are not homologous to any other parvulin-like PPIase[169]. So far, the function of these flanking regions remained unclear, and no structural basis for PrsA function has been reported.

2 The Dynamic Organization of Fungal Acetyl-CoA Carboxylase

This research was originally published in Nature Communications.

Moritz Hunkeler*, Edward Stuttfeld*, Anna Hagmann, Stefan Imseng, Timm Maier.

The Dynamic Organization of Fungal Acetyl-CoA Carboxylase.

Nat Commun. 2016; **7:** 11196.

2.1 Abstract

Acetyl-CoA carboxylases (ACCs) catalyze the committed step in fatty acid biosynthesis, the ATP-dependent carboxylation of acetyl-CoA to malonyl-CoA. They are important regulatory hubs for metabolic control and relevant drug targets for the treatment of the metabolic syndrome and cancer. Eukaryotic ACCs are single-chain multienzymes characterized by a large, non-catalytic central domain (CD), whose role in ACC regulation remains poorly characterized. The crystal structure of the yeast ACC CD reveals a unique dynamic four-domain organization. A regulatory loop, which is phosphorylated at the key functional phosphorylation site of fungal ACC, wedges into a crevice between two domains of CD. Combining the yeast CD structure with intermediate and low-resolution data of larger fragments up to intact ACCs provides a comprehensive characterization of the dynamic fungal ACC architecture. In contrast to related carboxylases, large-scale conformational changes are required for substrate turnover, and are mediated by the CD under phosphorylation control.

2.2 Introduction

Biotin-dependent acetyl-CoA carboxylases (ACCs) are essential enzymes that catalyze the ATP-dependent carboxylation of acetyl-CoA to malonyl-CoA. This reaction provides the committed activated substrate for the biosynthesis of fatty acids via fatty acid synthase [45,89]. By catalyzing this rate-limiting step in fatty acid biosynthesis, ACC plays a key role in anabolic metabolism. ACC inhibition and knock-out studies show the potential of targeting ACC for treatment of the metabolic syndrome [90,97,98]. Furthermore, elevated ACC activity is observed in malignant tumors [178,179]. A direct link between ACC and cancer is provided by cancer-associated mutations in the breast cancer susceptibility gene 1 (BRCA1), which relieve inhibitory interactions of BRCA1 with ACC [129,130]. Thus, ACC is a relevant drug target for type 2 diabetes and cancer [44,180,181]. Microbial ACCs are also the principal target of the antifungal and antibiotic compound Soraphen A [87,133].

The principal functional protein components of ACCs have been described already in the late 1960's for *Escherichia coli* (*E. coli*) ACC [182,183]: Biotin carboxylase (BC) catalyzes the ATP-dependent carboxylation of a biotin moiety, which is covalently linked to the biotin carboxyl carrier protein (BCCP). Carboxyltransferase (CT) transfers the activated carboxyl group from carboxy-biotin to acetyl-CoA to yield malonyl-CoA. Prokaryotic ACCs are transient assemblies of individual BC, CT and BCCP subunits [184]. Eukaryotic ACCs, instead, are multienzymes, which integrate all functional components into a single polypeptide chain of approx. 2300 amino acids [45]. Human ACC occurs in two closely related isoforms, ACC1 and 2, located in the cytosol and at the outer mitochondrial membrane, respectively [185,186]. In addition to the canonical ACC components, eukaryotic ACCs contain two non-catalytic regions, the large central domain (CD) and the BC-CT interaction domain (BT). The CD comprises one third of the protein and is a unique feature of eukaryotic ACCs without homologs in other proteins. The function of this domain remains poorly characterized, although phosphorylation of several serine residues in the CD regulates ACC activity [80,93,187]. The BT domain has been visualized in bacterial carboxylases, where it mediates contacts between α- and β-subunits [58,59].

Structural studies on the functional architecture of intact ACCs have been hindered by their huge size and pronounced dynamics, as well as the transient

assembly mode of bacterial ACCs. However, crystal structures of individual components or domains from prokaryotic and eukaryotic ACCs, respectively, have been solved [79,88,188-190]. The structure determination of the holoenzymes of bacterial biotin-dependent carboxylases, which lack the characteristic CD, such as the pyruvate carboxylase (PC) [60], propionyl-CoA carboxylase (PCC) [59], 3-methyl-crotonyl-CoA carboxylase (MCC) [58] and a long-chain acyl-CoA carboxylase (LCC) [86] revealed strikingly divergent architectures despite a general conservation of all functional components. In these structures, the BC and CT active sites are at distances between 40 Å and 80 Å, such that substrate transfer could be mediated solely by the mobility of the flexibly tethered BCCP.

Human ACC1 is regulated allosterically, via specific protein-protein interactions, and by reversible phosphorylation. Dynamic polymerization of human ACC1 is linked to increased activity and is regulated allosterically by the activator citrate and the inhibitor palmitate [93,187] or by binding of the small protein MIG-12 [84]. Human ACC1 is further regulated by specific phosphorylation-dependent binding of BRCA1 to Ser1263 in the CD. BRCA1 binds only to the phosphorylated form of ACC1 and prevents ACC activation by phosphatase-mediated dephosphorylation [129,191]. Furthermore, phosphorylation by AMP-activated protein kinase (AMPK) and cAMP-dependent protein kinase (PKA) leads to a decrease in ACC1 activity. AMPK phosphorylates ACC1 *in vitro* at Ser80, Ser1201, and Ser1216 and PKA at Ser78 and Ser1201. However, regulatory effects on ACC1 activity are mainly mediated by phosphorylation of Ser80 and Ser1201, respectively [126,127]. Phosphorylated Ser80, which is highly conserved in higher eukaryotes, presumably binds into the Soraphen A binding pocket [128]. The regulatory Ser1201 shows only moderate conservation across higher eukaryotes, while the phosphorylated Ser1216 is highly conserved across all eukaryotes. However, no effect of Ser1216 phosphorylation on ACC activity has been reported in higher eukaryotes.

For fungal ACC, neither spontaneous nor inducible polymerization has been detected despite considerable sequence conservation to human ACC1. The BRCA1-interacting phosphoserine position is not conserved in fungal ACC, and no other phospho-dependent protein-protein interactions of fungal ACC have been described. In yeast ACC, phosphorylation sites have been identified at Ser2, Ser735, Ser1148, Ser1157 and Ser1162. Of these, only Ser1157 is highly conserved in fungal ACC and aligns to Ser1216 in human ACC1. Its

phosphorylation by the AMPK homologue Snf1 results in strongly reduced ACC activity [80,131,132].

Despite the outstanding relevance of ACC in primary metabolism and disease, the dynamic organization and regulation of the giant eukaryotic, and in particular fungal ACC, remain poorly characterized. Here, we provide the structure of *Sce*ACC CD, intermediate- and low-resolution structures of human (*Hsa*) ACC CD and larger fragments of fungal ACC from *Chaetomium thermophilum* (*Cth*) (Figure 2.1a). Integrating these data with small-angle X-ray scattering (SAXS) and electron microscopy (EM) observations yields a comprehensive representation of the dynamic structure and regulation of fungal ACC.

2.3 Results

2.3.1 The organization of the yeast ACC central domain

First, we focused on structure determination of the 82 kDa CD. The crystal structure of the CD of *Sce*ACC (*Sce*CD) was determined at 3.0 Å resolution by experimental phasing and refined to R_{work}/R_{free}= 0.20/0.25 (Supplemental Table 2.1). The overall extent of the *Sce*CD is 70 by 75 Å (Figure 2.1b, Supplemental Figure 2.1a, b), and the attachment points of the N-terminal 26-residue linker to the BCCP domain and the C-terminal CT domain are separated by 46 Å (the N- and C-termini are indicated with spheres in Figure 2.1b). *Sce*CD comprises four distinct domains, an N-terminal α-helical domain (CD_N), and a central four-helix bundle linker domain (CD_L), followed by two α-β-fold C-terminal domains (CD_{C1}/CD_{C2}). CD_N adopts a letter C shape, where one of the ends is a regular four-helix bundle (Nα3-6), the other end is a helical hairpin (Nα8,9) and the bridging region comprises six helices (Nα1,2,7,10-12). CD_L is composed of a small, irregular four-helix bundle (Lα1-4) and tightly interacts with the open face of CD_{C1} via an interface of 1300 $Å^2$ involving helices Lα3 and Lα4. CD_L does not interact with CD_N apart from the covalent linkage and forms only a small contact to CD_{C2} via a loop between Lα2/α3 and the N-terminal end of Lα1 with an interface area of 400 $Å^2$. CD_{C1}/CD_{C2} share a common fold; they are composed of six-stranded β-sheets flanked on one side by two long, bent helices inserted between strands β3/β4 and β4/β5. CD_{C2} is extended at its C-terminus by an additional β-strand and an irregular β-hairpin.

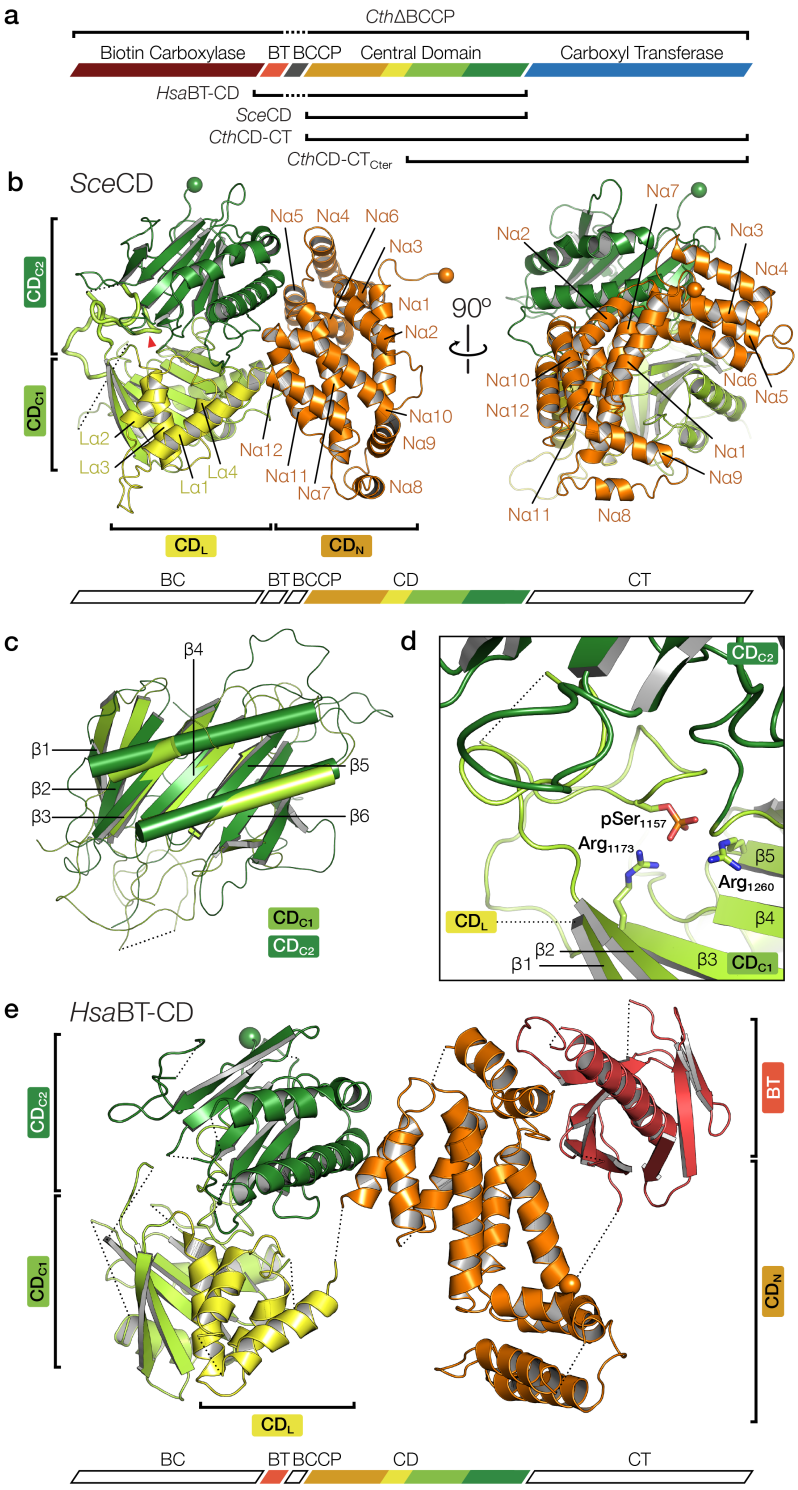


Figure 2.1 The phosphorylated central domain of yeast ACC.
(**a**) Schematic overview of the domain organization of eukaryotic ACCs. Crystallized constructs are indicated. (**b**) Cartoon representation of the *Sce*CD crystal structure. CD_N is linked by a four-helix bundle (CD_L) to two α-β fold domains (CD_{C1} and CD_{C2}). The regulatory loop is shown as bold cartoon and the phosphorylated Ser1157 is marked by a red triangle. The N- and C-termini are indicted by spheres. (**c**) Superposition of CD_{C1} and CD_{C2} reveals highly conserved folds. (**d**) The regulatory loop with the phosphorylated Ser1157 is bound into a crevice between CD_{C1} and CD_{C2}, the conserved residues Arg1173 and Arg1260 coordinate the phosphoryl-group. (**e**) Structural overview of *Hsa*BT-CD. The attachment points to the N-terminal BCCP domain and the C-terminal CT domain are indicated with spheres. All coloring is according to scheme (a).

Based on an rmsd of main chain atom positions of 2.2 Å, CD_{C1}/CD_{C2} are structurally more closely related to each other than to any other protein (Figure 2.1c); they may thus have evolved by duplication. Close structural homologs could not be found for the CD_N or the CD_C domains.

2.3.2 A regulatory loop mediates interdomain interactions

To define the functional state of insect-cell expressed ACC variants, we employed mass spectrometry (MS) for phosphorylation site detection. In insect-cell expressed full-length *Sce*ACC, the highly conserved Ser1157 is the only fully occupied phosphorylation site with functional relevance in *S. cerevisiae*. Additional phosphorylation was detected for Ser2101 and Tyr2179, however these sites are neither conserved across fungal ACC nor natively phosphorylated in yeast. MS analysis of dissolved crystals confirmed the phosphorylated state of Ser1157 also in *Sce*CD crystals. The *Sce*CD structure thus authentically represents the state of *Sce*ACC, where the enzyme is inhibited by Snf1-dependent phosphorylation.

In the *Sce*CD crystal structure, the phosphorylated Ser1157 resides in a regulatory 36-amino acid loop between strands β2 and β3 of CD_{C1} (Figure 2.1b, d), which contains two additional less conserved phosphorylation sites (Ser1148, Ser1162) confirmed in yeast [192], but not occupied here. This regulatory loop wedges between the CD_{C1} and CD_{C2} domains and provides the largest contribution to the interdomain interface. The N-terminal region of the regulatory loop also directly contacts the C-terminal region of CD_{C2} leading into CT. Phosphoserine 1157 is tightly bound by two highly conserved arginines (Arg1173, Arg1260) of CD_{C1} (Figure 2.1d). Already the binding of phosphorylated Ser1157 apparently stabilizes the regulatory loop conformation, the accessory phosphorylation sites Ser1148 and Ser1162 in the same loop may further modulate the strength of interaction between the regulatory loop and the CD_{C1} and CD_{C2} domains. Phosphorylation of the regulatory loop thus determines interdomain interactions of CD_{C1} and CD_{C2} suggesting that it may exert its regulatory function by modifying the overall structure and dynamics of the CD.

The functional role of Ser1157 was confirmed by an activity assay based on the incorporation of radioactive carbonate into acid non-volatile material [193]. Phosphorylated *Sce*ACC shows only residual activity (k_{cat} = 0.4 ± 0.2 s^{-1}), which

increases 16-fold (k_{cat} = 6.5 ± 0.3 s^{-1}) after dephosphorylation with λ protein phosphatase. The values obtained for dephosphorylated *Sce*ACC are comparable to earlier measurements of non-phosphorylated yeast ACC expressed in *E. coli* [194].

2.3.3 The variable CD is conserved between yeast and human

To compare the organization of fungal and human ACC CD, we determined the structure of a human ACC1 fragment that comprises the BT and CD domains (*Hsa*BT-CD), but lacks the mobile BCCP in between (Figure 2.1a). An experimentally-phased map was obtained at 3.7 Å resolution for a cadmium-derivatized crystal and was interpreted by a poly-alanine model (Figure 2.1e, Tab.1). Each of the four CD domains in *Hsa*BT-CD individually resembles the corresponding *Sce*CD domain; however, human and yeast CDs exhibit distinct overall structures. In agreement with their tight interaction in *Sce*CD, the relative spatial arrangement of CD_{C1} and CD_L is preserved in *Hsa*BT-CD, but the human CD_L/CD_{C1} didomain is tilted by 30° based on a superposition of human and yeast CD_{C2} (Supplemental Figure 2.1c). As a result, the N-terminus of CD_L at helix Lα1, which connects to CD_N, is shifted by 12 Å. Remarkably, CD_N of *Hsa*BT-CD adopts a completely different orientation compared to *Sce*CD. With CD_L/CD_{C1} superposed, CD_N in *Hsa*BT-CD is rotated by 160° around a hinge at the connection of CD_N/CD_L (Supplemental Figure 2.1d). This rotation displaces the N-terminus of CD_N in *Hsa*BT-CD by 51 Å compared to *Sce*CD resulting in a separation of the attachment points of the N-terminal linker to the BCCP domain and the C-terminal CT domain by 67 Å (the attachment points are indicated with spheres in Figure 2.1e). The BT domain of *Hsa*BT-CD consists of a helix that is surrounded at its N-terminus by an anti-parallel eight-stranded β-barrel. It resembles the BT of PCC [59], only the four C-terminal strands of the β-barrel are slightly tilted.

Based on MS analysis, insect-cell expressed human full-length ACC Ser80 shows the highest degree of phosphorylation (90%). Ser29 and Ser1263, implicated in insulin-dependent phosphorylation and BRCA1 binding, respectively, are phosphorylated at intermediate levels (40%). The highly conserved Ser1216 (corresponding to *S. cerevisiae* Ser1157), as well as Ser1201, both in the regulatory loop discussed above, are not phosphorylated. However,

residual phosphorylation levels were detected for Ser1204 (7%) and Ser1218 (7%) in the same loop. MS analysis of the *Hsa*BT-CD crystallization sample reveals partial proteolytic digestion of the regulatory loop. Accordingly, most of this loop is not represented in the *Hsa*BT-CD crystal structure. The absence of the regulatory loop might be linked to the less restrained interface of CD_{C2} and CD_L/CD_{C1} and altered relative orientations of these domains. Besides the regulatory loop, also the phosphopeptide target region for BRCA1 interaction is not resolved presumably due to pronounced flexibility.

At the level of isolated yeast and human CD, the structural analysis indicates the presence of at least two hinges, one with large-scale flexibility at the CD_N/CD_L connection, and one with tunable plasticity between CD_{C2} and CD_L/CD_{C1}, plausibly affected by phosphorylation in the regulatory loop region.

2.3.4 The integration of CD into the fungal ACC multienzyme

To further obtain insights into the functional architecture of fungal ACC, we characterized larger multi-domain fragments up to the intact enzymes. Using molecular replacement based on fungal ACC BC, CD and CT models, we obtained structures of a variant comprising *Cth*CT and CD_{C1}/CD_{C2} in two crystal forms at resolutions of 3.6 Å and 4.5 Å (*Cth*CD-$CT_{Cter1/2}$), respectively, as well as of a *Cth*CT linked to the entire CD at 7.2 Å resolution (*Cth*CD-CT) (Figure 2.1a, 2, Tab. 1). No crystals diffracting to sufficient resolution were obtained for larger BC-containing fragments, or for full-length *Cth* or *Sce*ACC. To improve crystallizibility, we generated ΔBCCP variants of full-length ACC, which based on SAXS analysis preserve properties of intact ACC (Supplemental Figure 2.2a-c). For *Cth*ΔBCCP, crystals diffracting to 8.4 Å resolution were obtained. However, molecular replacement did not reveal a unique positioning of the BC domain. Due to the limited resolution the discussion of structures of *Cth*CD-CT and *Cth*ΔBCCP is restricted to the analysis of domain localization. Still, these structures contribute considerably to the visualization of an intrinsically dynamic fungal ACC.

In all these crystal structures, the CT domains build a canonical head-to-tail dimer [190], with active sites formed by contributions from both protomers (Figure 2.2, Supplemental Figure 2.3a). The connection of CD and CT is provided by a 10-residue peptide stretch, which links the N-terminus of CT to the irregular β-

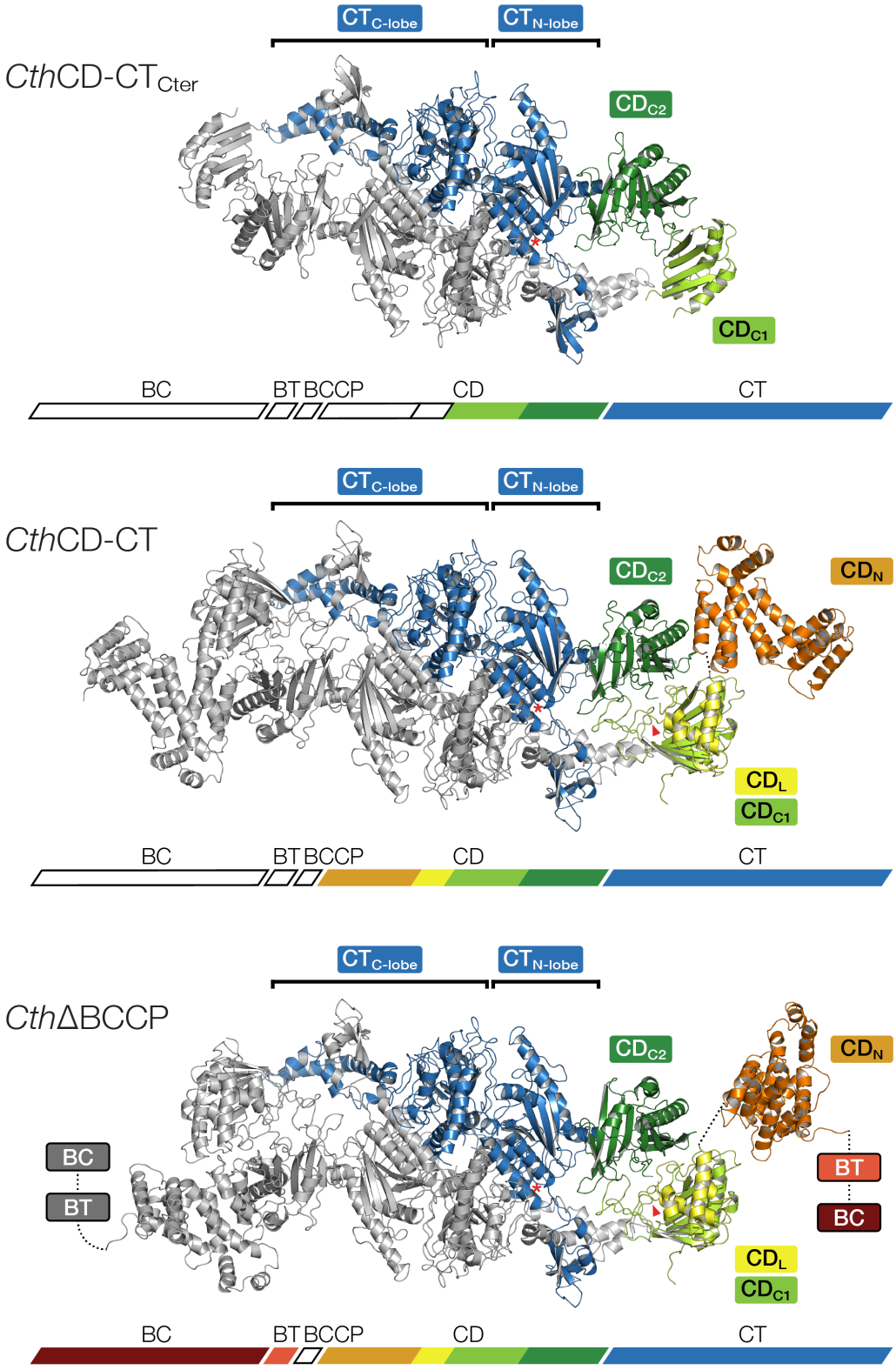


Figure 2.2 Architecture of the CD-CT core of fungal ACC.
Cartoon representation of crystal structures of multi-domain constructs of *Cth*ACC. One protomer is shown in color and one in gray. Individual domains are labeled, the active site of CT and the position of the conserved phosphoserine site (Ser1157 in *S. cerevisiae* ACC) are indicated by an asterisk and a triangle, respectively.

hairpin/β-strand extension of CD_{C2} (Supplemental Figure 2.3b). The connecting region is remarkably similar in isolated CD and *Cth*CD-CT_{Cter} structures, indicating inherent conformational stability. CD/CT contacts are only formed in direct vicinity of the covalent linkage and involve the β-hairpin extension of CD_{C2} as well as the loop between strands β2/β3 of the CT N-lobe, which contains a conserved RxxGxN motif. The neighboring loop on the CT side (between CT β1/β2) is displaced by 2.5 Å compared to isolated CT structures (Supplemental Figure 2.3c). Based on an interface area of around 600 $Å^2$ and its edge-to-edge connection characteristics, the interface between CT and CD might be classified as conformationally variable. Indeed, the comparison of the positioning of eight instances of the C-terminal part of CD relative to CT in crystal structures determined here, reveals flexible interdomain linking (Figure 2.3a): The CD_{C2}/CT interface acts as a true hinge with observed rotation up to 16°, which results in a translocation of the distal end of CD_{C2} by 8 Å.

The interface between CD_{C2} and CD_L/CD_{C1}, which is mediated by the phosphorylated regulatory loop in the *Sce*CD structure, is less variable than the CD-CT junction, and permits only limited rotation and tilting (Figure 2.3b). Analysis of the impact of phosphorylation on the interface between CD_{C2} and CD_L/CD_{C1} in *Cth*ACC variant structures is precluded by the limited crystallographic resolution. However, MS analysis of *Cth*CD-CT and CthΔBCCP constructs revealed between 60 and 70% phosphorylation of Ser1170 (corresponding to *Sce*ACC Ser1157). The CD_N domain positioning relative to CD_L/CD_{C1} is highly variable with three main orientations observed in the structures of *Sce*CD and the larger *Cth*ACC fragments: CD_N tilts, resulting in a displacement of its N-terminus by 23 Å (Figure 2.4a, observed in both protomers of *Cth*CD-CT and one protomer of *Cth*ΔBCCP). In addition, CD_N can rotate around hinges in the connection between CD_N/CD_L by 70° (Figure 2.4b, observed in the second protomer of *Cth*ΔBCCP) and 160° (Figure 2.4c, observed in *Sce*CD) leading to displacement of the anchor site for the BCCP linker by up to 33 Å and 40 Å, respectively.

Conformational variability in the CD thus contributes considerably to variations in the spacing between the BC and CT domains, and may extend to distance variations beyond the mobility range of the flexibly tethered BCCP. Based on the occurrence of the distinct conformations between fungal and human ACC fragments, they may well represent general conformational states present in all eukaryotic ACCs.

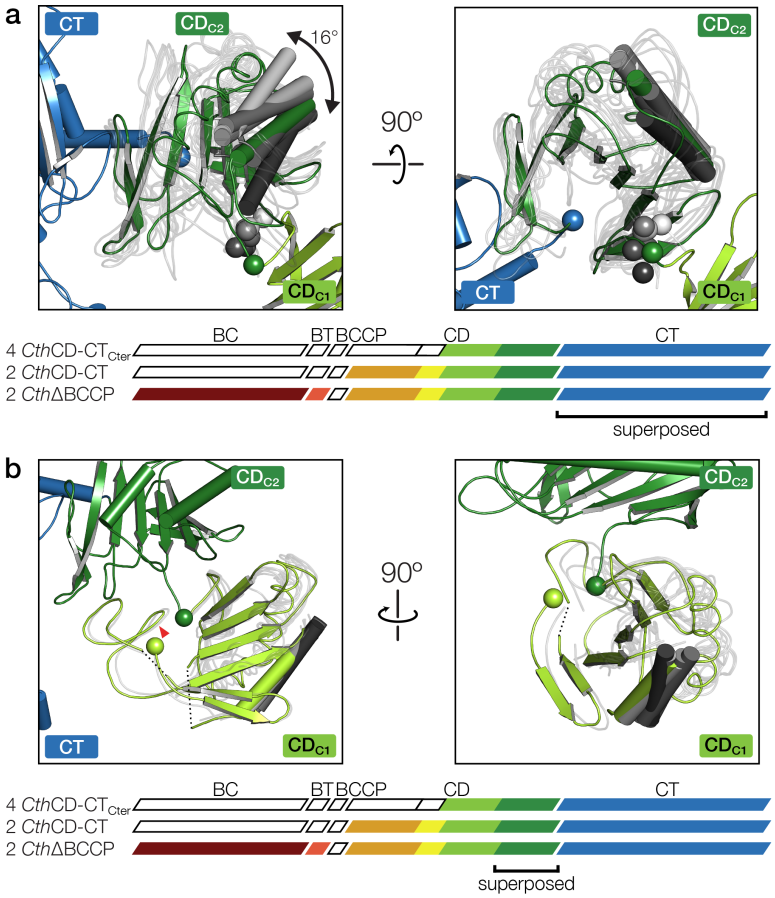


Figure 2.3 Structural variability of the CD-CT connection in fungal ACC.
(**a**) Hinge properties of the CD-CT connection analyzed by a CT-based superposition of eight instances of the CD_{C1}/CD_{C2}-CT segment. For clarity, only one protomer *Cth*CD-CT_{Cter1} is shown in full color as reference. For other instances, CD_{C2} domains are shown in transparent tube representation with only one helix each highlighted. The range of hinge bending is indicated and the connection points between CD_{C2} and CT (blue) as well as between CD_{C1} and CD_{C2} (green and gray) are marked as spheres. (**b**) The interdomain interface of CD_{C1} and CD_{C2} exhibits only limited plasticity. Representation as in (a), but the CD_{C1} and CD_{C2} are superposed based on CD_{C2}. One protomer of *Cth*ΔBCCP is shown in color, the CD_L domains are omitted for clarity and the position of the phosphorylated serine is indicated with a red triangle. The connection points from CD_{C1} to CD_{C2} and to CD_L are represented with green spheres.

2.3.5 Large-scale conformational variability of fungal ACC

To obtain a comprehensive view of fungal ACC dynamics in solution, we have employed SAXS and EM. SAXS analysis of *Cth*ACC agrees with a dimeric state and an elongated shape with a maximum extent of 350 Å (Supplemental Table 2.2). The smooth appearance of scattering curves and derived distance distributions might indicate substantial interdomain flexibility [195] (Supplemental Figure 2.2a-c). Direct observation of individual full-length C*th*ACC particles, according to MS results predominantly in a phosphorylated low activity state, in

negative stain electron microscopy reveals a large set of conformations from rod-like extended to U-shaped particles.

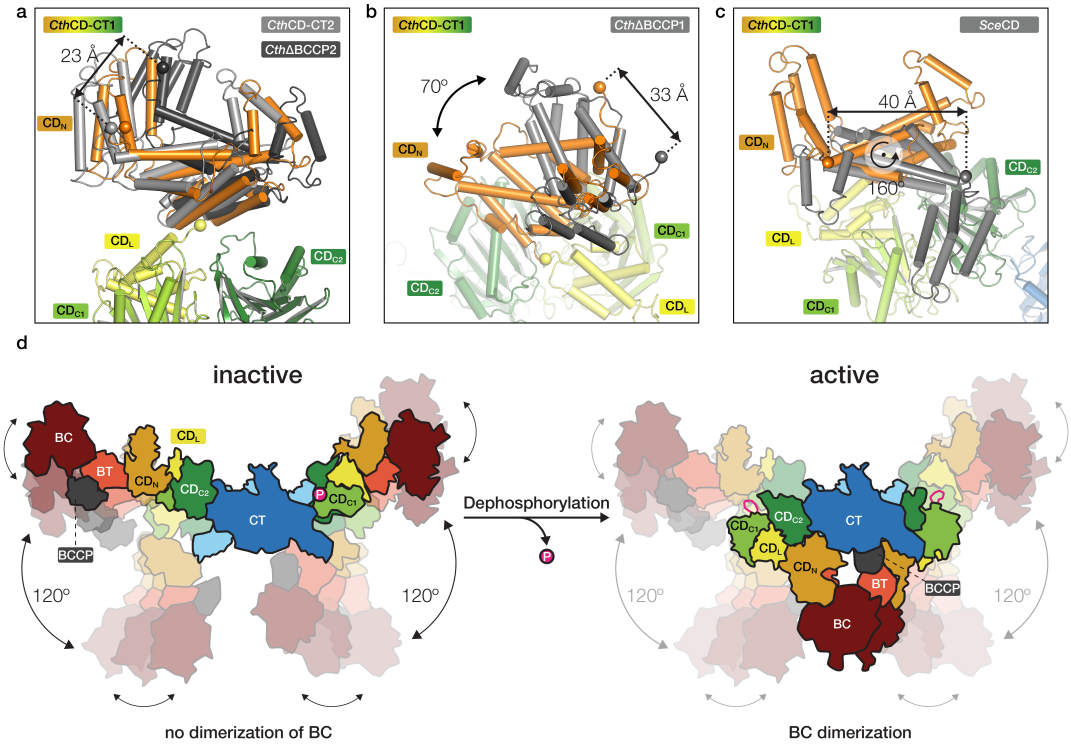


Figure 2.4 The conformational dynamics of fungal ACC.
(**a**)-(**c**) Large-scale conformational variability of the rotation of the CD_N domain relative to the CD_L/CD_{C1} domain. *Cth*CD-CT1 (in color) serves as reference, the compared structures (as indicated) are shown in gray. Domains other than CD_N and CD_L/CD_{C1} are omitted for clarity. The domains are labeled and the distances between the N-termini of CD_N (spheres) in the compared structures are indicated. (**d**) Schematic model of fungal ACC showing the intrinsic, regulated flexibility of CD in the phosphorylated, inhibited and the non-phosphorylated activated state. Flexibility of the CD_{C2}/CT and CD_N/CD_L hinges is illustrated by arrows. The Ser1157 phosphorylation site and the dephosphorylated regulatory loop are colored in magenta.

Class averages, obtained by maximum-likelihood-based 2D classification, are focused on the dimeric CT domain and the full BC-BCCP-CD domain of only one protomer, due to the non-coordinated motions of the lateral BC/CD regions relative to the CT dimer. They identify the connections between CD_N/CD_L and between CD_{C2}/CT as major contributors to conformational heterogeneity (Supplemental Figure 2.4a, b). The flexibility in the CD_{C2}/CT hinge appears substantially larger than the variations observed in the set of crystal structures. The BC domain is not completely disordered, but laterally attached to BT/CD_N in a generally conserved position, albeit with increased flexibility. Surprisingly, in both the linear and U-shaped conformations, the approximate distances between the BC and CT active sites would remain larger than 110 Å. These observed distances are considerably larger than in static structures of any other related

biotin-dependent carboxylase. Furthermore, based on an average length of the BCCP-CD linker in fungal ACC of 26 amino acids, mobility of the BCCP alone would not be sufficient to bridge the active sites of BC and CT. Consequently, increased flexibility or additional modes of conformational changes may be required for productive catalysis. The most relevant candidate site for mediating such additional flexibility and permitting an extended set of conformations is the CD_{C1}/CD_{C2} interface, which is rigidified by the Ser1157-phosphorylated regulatory loop as depicted in the *Sce*CD crystal structure.

2.4 Discussion

Altogether, the architecture of fungal ACC is based on the central dimeric CT domain (Figure 2.4d). The CD consists of four distinct subdomains and it acts as a tether from the CT to the mobile BCCP and an oriented BC domain. The CD has no direct role in substrate recognition or catalysis but contributes to the regulation of all eukaryotic ACCs. In higher eukaryotic ACCs, regulation via phosphorylation is achieved by combining the effects of phosphorylation at Ser80, Ser1201 and Ser1263. In fungal ACC however, Ser1157 in the regulatory loop of the CD is the only phosphorylation site that has been demonstrated to be both phosphorylated *in vivo* and involved in the regulation of ACC activity. In its phosphorylated state, the regulatory loop containing Ser1157 wedges between CD_{C1}/CD_{C2} and presumably limits the conformational freedom at this interdomain interface. However, flexibility at this hinge may be required for full ACC activity, as the distances between the BCCP anchor points and the active sites of BC and CT observed here are such large, that mobility of the BCCP alone is not sufficient for substrate transfer. The current data thus suggest that regulation of fungal ACC is mediated by controlling the dynamics of the unique CD, rather than directly affecting catalytic turnover at the active sites of BC and CT. A comparison between fungal and human ACC will help to further discriminate mechanistic differences that contribute to the extended control and polymerization of human ACC.

Most recently, a crystal structure of near full-length non-phosphorylated ACC from *S. cerevisae* (lacking only 21 N-terminal amino acids, here denoted as flACC) was published by Wei and Tong [194]. In flACC, the ACC dimer obeys two-fold symmetry and assembles in a triangular architecture with dimeric BC domains (Supplemental Figure 2.5a). In their study, mutational data indicate a requirement for BC dimerization for catalytic activity. The transition from the elongated open

shape, observed in our experiments, towards a compact triangular shape is based on an intricate interplay of several hinge-bending motions in the CD (Figure 2.4d). Comparison of flACC to our *Cth*ΔBCCP structure reveals the CD_{C2}/CT hinge as a major contributor to conformational flexibility (Supplemental Figure 2.5b, c). In flACC, CD_{C2} rotates approximately 120° with respect to the CT domain. A second hinge can be identified between CD_{C1}/CD_{C2}. Based on a superposition of CD_{C2}, CD_{C1} of the phosphorylated *Sce*CD is rotated by 30° relative to CD_{C1} of the non-phosphorylated flACC (Supplemental Figure 2.5d), similar to what we have observed for the non-phosphorylated *Hsa*BT-CD (Supplemental Figure 2.1d). When inspecting all individual protomer and fragment structures in their study, Wei & Tong also identify the CD_N/CD_{C1} connection as a highly flexible hinge, in agreement with our observations.

The only bona-fide regulatory phophorylation site of fungal ACC in the regulatory loop is directly participating in CD_{C1}/CD_{C2} domain interactions and thus stabilizes the hinge conformation. In flACC, the regulatory loop is mostly disordered, illustrating the increased flexibility due to the absence of the phosphoryl-group. Only in three out of eight observed protomers a short peptide stretch (including Ser1157) was modeled, and in those instances the Ser1157 residue is located at a distance of 14 – 20 Å away from the location of the phosphorylated serine observed here for a superposition based on either CD_{C1} or CD_{C2}. Applying the conformation of the CD_{C1}/CD_{C2} hinge observed in *Sce*CD onto flACC leads to CD_N sterically clashing with CD_{C2} and BT/CD_N clashing with CT (Supplemental Figure 2.6a, b). Thus, in accordance with the results presented here, phosphorylation of Ser1157 in *Sce*ACC most likely limits flexibility in the CD_{C1}/CD_{C2} hinge such that activation through BC dimerization is not possible (Figure 2.4d), which however does not exclude intermolecular dimerization. In addition, EM micrographs of phosphorylated and dephosphorylated *Sce*ACC display for both samples mainly elongated and U-shaped conformations and reveal no apparent differences in particle shape distributions (Supplemental Figure 2.7). This implicates that the triangular shape with dim-eric BC domains has a low population also in the active form, even though a biasing influence of grid preparation cannot be excluded completely.

Large-scale conformational variability has also been observed in most other carrier protein based multienzymes, including polyketide [196] and fatty acid synthases (with the exception of fungal-type fatty acid synthases) [25,26], non-ribosomal peptide synthetases [197] and the pyruvate dehydrogenase complexes

[198], although based on completely different architectures. Together, this structural information suggests that variable carrier protein tethering is not sufficient for efficient substrate transfer and catalysis in any of these systems. The determination of a set of crystal structures of *Sce*ACC in two states, unphosphorylated [194] and phosphorylated at the major regulatory site Ser1157, provides a unique depiction of multienzyme regulation by post-translational modification (Figure 2.4d). The phosphorylated regulatory loop binds to an allosteric site at the interface of two non-catalytic domains and restricts conformational freedom at several hinges in the dynamic ACC. It disfavors the adoption of a rare, compact conformation, in which intramolecular dimerization of the BC domains [194] results in catalytic turnover. The regulation of activity thus results from restrained large-scale conformational dynamics rather than a direct or indirect influence on active site structure. To our best knowledge, ACC is the first multienzyme for which such phosphorylation-dependent mechanical control mechanism has been visualized. However, the example of ACC now demonstrates the possibility of regulating activity by controlled dynamics of non-enzymatic linker regions also in other families of carrier-dependent multienzymes. Understanding such structural and dynamic constraints imposed by scaffolding and linking in carrier protein based multienzyme systems is a critical prerequisite for engineering of efficient biosynthetic assembly lines.

2.5 Methods

2.5.1 Protein Expression & Purification

All proteins were expressed in the Baculovirus Expression Vector System. The MultiBac™ insect cell expression plasmid pACEBACI (Geneva Biotech) was modified to host a GATEWAY® (LifeTechnologies) cassette with an N-terminal 10xHis-tag, named pAB1GW-NH10 hereafter. Full-length *Hsa*ACC (Genebank accession #U19822), *Sce*ACC (#NC_001146), *Cth*ACC (#XM_006692575) were cloned into pAB1GW-NH10 using GATEWAY® according to manufacturers manual. Truncated variants were constructed by PCR amplification, digestion of the template DNA with DpnI, phosphorylation of the PCR-product and religation of the linear fragment to a circular plasmid. The following constructs were used for this study: *Sce*ACC (1–2233), *Cth*ACC (1–2297), *Cth*ΔBCCP (1–2297, Δ700–765), *Cth*CD-CT (788–2297), *Cth*CD-CT_{Cter} (1114–2297), *Sce*CD (768–1494), *Hsa*BT-CD (622–1584, Δ753–818). Bacmid and virus production was carried out according to MultiBac™ instructions [199]. Baculovirus generation and amplification as well as protein expression were performed in Sf21 cells (Expression Systems) in Insect-Xpress medium (Lonza). The cells were harvested between 68-96h post infection by centrifugation and stored at -80 °C until being processed.

Cells were lysed by sonication and the lysate was cleared by ultracentrifugation. Soluble protein was purified using Ni-NTA (Genscript) and size exclusion chromatography (Superose 6, GE Healthcare). The affinity tag was removed by TEV-cleavage overnight at 4°C. TEV-protease and uncleaved protein were removed by orthogonal Ni-NTA purification prior to size exclusion chromatography. *Sce*ACC, *Cth*ACC and *Cth*ΔBCCP were further purified by high-resolution anion exchange chromatography prior to size exclusion chromatography. Purified *Sce*CD, *Cth*CD-CT_{Cter}, *Cth*CD-CT, *Cth*ΔBCCP, *Sce*ACC were concentrated to 10 mg ml^{-1} in 30 mM MOPS pH 7, 200 mM ammonium sulfate, 5% glycerol and 10 mM dithiothreitol and purified *Hsa*BT-CD was concentrated to 20 mg ml^{-1} in 20 mM bicine pH 8.0, 200 mM NaCl, 5% glycerol, and 5 mM TCEP. Proteins were used directly or were stored at –80°C after flash-freezing in liquid nitrogen.

2.5.2 Protein Crystallization

All crystallization experiments were conducted using sitting drop vapour diffusion. *Sce*CD crystals were grown at 19 °C by mixing protein and reservoir solution (0.1 M BisTrisPropane pH 6.5, 0.05 – 0.2 M di-sodium malonate, 20 – 30% polyethylene glycol (PEG) 3350, 10 mM trimethylamine or 2% benzamidine) in a 1:1 or 2:1 ratio. Crystals appeared after several days and continued to grow for 20 to 200 days. Crystals were cryo-protected by short incubation in mother liquor supplemented with 22% ethylene glycol and flash-cooled in liquid nitrogen. For heavy metal derivatization the crystals were incubated in stabilization solution supplemented with 1 mM Thimerosal or 10 mM $EuCl_2$, respectively, and then backsoaked for 15 seconds in stabilization solution without heavy metal.

Initial crystals of *Hsa*BT-CD grew in 0.1 M Tris pH 8.5, 0.35 M tri-potassium citrate and 2 – 3.5% PEG10000 at 19 °C. After several rounds of optimization good quality diffraction crystals were obtained at 19 °C in 0.1M MES pH 6, 0.25 – 0.35 M tri-potassium citrate, 2 – 5% PEG10000, 0.01 – 0.04 M cadmium chloride. The protein drop contained a 1:1 ratio of protein and reservoir solution. Crystals grew immediately and stopped growing after 3 days. They were dehydrated and cryoprotected in several steps in artificial mother liquor containing incrementally increasing concentrations of tri-potassium citrate, PEG10000 and ethylene glycol and then flash-cooled in liquid nitrogen. The final solution was composed of 0.1 M MES pH6, 0.5 M tri-potassium citrate, 6.75% PEG10000, 0.01 M cadmium chloride.

*Cth*CD-CT_{Cter} crystals were grown at 19°C by mixing protein and reservoir solution (0.1 M Hepes pH 7.5, 2 – 7% Tacsimate pH 7, 7.5 – 15% PEGMME5000) in a 1:1 ratio. Crystals appeared after several days and continued to grow for up to two weeks. Crystals were cryo-protected by short incubation in mother liquor supplemented with 22% ethylene glycol.

*Cth*CD-CT ACC crystals were grown at 19 °C by mixing protein and reservoir solution (0.1 M Bicine pH 8.5 – 9.5, 4 – 8% PEG8000) in a 1:1 or 1:2 ratio. Crystals grew 8 to 10 days and were cryo-protected by short incubation in mother liquor supplemented with 22% ethylene glycol before flash-cooling in liquid nitrogen.

*Cth*ΔBCCP ACC crystals were grown at 19 °C by mixing protein and reservoir solution (0.1 M morpheus buffer 3, 7 – 12% Morpheus ethylene glycols mix, 8 – 12% PEG4000, 17 – 23% glycerol) in a 1:1 or 1:2 ratio. Crystals grew up to three

weeks and were cryo-protected in reservoir solution before flash cooling in liquid nitrogen.

2.5.3 Structure Determination and Analysis of Phosphorylation

All X-ray diffraction data were collected at beamlines X06SA (PXI) and X06DA (PXIII) at the Swiss Light Source (SLS, Paul Scherrer Institute, Villigen, Switzerland) equipped with PILATUS detectors, respectively. The wavelength of data collection was 1.000 Å for native crystals, and 1.527 Å and 1.907 Å for crystals derivatized with europium and cadmium, respectively. Raw data were processed using XDS [200]. Molecular replacement was carried out using Phaser 2.5.7 and 2.6.0, density modification was done using Parrot [201,202] and resolve, multi crystal averaging [203] was carried out using phenix. All model building was conducted using Coot [204] and figures were prepared using PyMOL (Schrödinger LLC).

Diffraction of initial *Sce*CD crystals in space group $P4_32_12$ with unit cell dimensions of a = b = 110.3 Å and c = 131.7 Å was limited to 3.5 Å. The resolution was improved to 3 Å by addition of trimethylamine or benzamidine to the reservoir solution without significant changes in unit cell dimensions. Crystals derivatized with thimerosal and europium were used for initial SAD phase determination using the SHELXC/D package [205]. Two mercury and four europium sites were located and an initial model was placed in the resulting maps. Since crystals derivatized with europium were slightly non-isomorphous with a c-axis length of 127 Å, multi crystal averaging was used for density modification and provided directly interpretable maps. Iterative cycles of model building and refinement in Buster (version 2.10.2; Global Phasing Ltd.) converged at R_{work}/R_{free} of 0.20/0.25. The final model lacks the disordered N-terminus (amino acids 768 – 789), an extended loop in the CD_{C1} domain (1204 – 1215) and a short stretch (1147 – 1152) preceding the regulatory loop. Based on temperature factor analysis, the start and end of the regulatory loop show higher disorder than the region around the interacting phosphoserine 1157. MS analysis of dissolved crystals detected quantitative phosphorylation of the regulatory Ser1157, as also found for full length *Sce*ACC, and additionally albeit with much lower occurrence, phosphorylation of Ser790, Ser1137, Ser1148 and Ser1159. A modeled phosphoryl-position for Ser1159 could overlap with the one of Ser1157, and might be represented in the crystal. For all other phosphorylation sites no

difference density could be observed, probably due to very low occupancy. PDBeFold [206] was used to search for structural homologues. The thresholds for lowest acceptable percentage of matched secondary structure elements were 70% for the search query and 20% for the result.

Initial *Hsa*BT-CD crystals were obtained in space group $I4_122$ with a = b = 240.1 Å and c = 768.9 Å and diffracted to only 7.5 Å. Optimized and dehydrated crystals also belonged to space group $I4_122$ but with unit cell parameters a = b = 267.3 Å and c = 210.6 Å and diffracted up to a resolution of 3.7 Å. Phase information was obtained from SAD based on bound cadmium ions from the crystallization condition. Six cadmium positions were located in a 4.0 Å resolution dataset at 1.9 Å wavelength using SHELXC/D [205] via the HKL2MAP interface [207]. Density modification and phasing based on this anomalous dataset, a 3.7 Å resolution data set at 1.0 Å wavelength, and additional non-isomorphous lower resolution datasets led to a high-quality electron density map. At the intermediate resolution obtained, the map was interpreted by a poly-alanine model, which was guided by predicted secondary structure as well as sequence and structural alignment with *Sce*CD. The final model refines against experimental data with with R_{work}/R_{free} of 0.38/0.41, as expected for a poly-alanine model [203]. Two *Hsa*BT-CD monomers are packed in the asymmetric unit via the CD_N and BT domains. Density on top of the β-barrel of one BT most likely representing parts of the BT-CD linker guided the assignment of this BT to its linked CD partner domain. This BT-to-CD assignment was further supported by the analysis of an additional lower resolution crystal form. Cadmium ions were found to participate in crystal packing.

In *Hsa*ACC. phosphorylation at regulatory sites was detected as provided in the main text. No phosphorylation was detected for other phosphosites previously identified in large-scale phosphoproteomics studies, namely serines 5, 23, 25, 48, 53, 78, 488, 786, 1273 [208-210].

Two different crystal forms were obtained for *Cth*CD-CT_{Cter} (denoted as *Cth*CD-CT_{Cter1} and *Cth*CD-CT_{Cter1}), diffracting to 3.6 Å and 4.5 Å, respectively. Both forms packed in space group $P2_12_12_1$ with unit cell constants of a = 97.7 Å, b = 165.3 Å and c = 219.2 Å or a = 100.2 Å, b = 153.5 Å and c = 249.2 Å, respectively. Phases were determined by molecular replacement using a homology model based on *Sce*CT (pdb 1od2) as search model in Phaser [202,211,212], multi crystal averaging was applied in density modification. The CT domain was rebuilt and an

initial homology model based on the *Sce*CD structure was fitted into difference density for *Cth*CD-CT_{Cter1}. Iterative cycles of rebuilding and refinement in Buster converged at R_{work}/R_{free} of 0.20/0.25. The refined CD fragment served as a starting model for rebuilding *Cth*CD-CT_{Cter2} at lower resolution. Coordinate refinement in Buster was additionally guided by reference model restraints and converged at R_{work}/R_{free} of 0.26/0.30. Residues 1114 – 1188, 1213 – 1252, 1380 – 1385 and 2188 – 2195 were disordered in both crystal forms and are not included in the models. Conservation was mapped onto the *Cth*CD-CT_{Cter1} crystal structure using al2co [213] based on a sequence alignment of 367 fungal ACC sequences calculated by Clustal Omega [214]. MS analysis of purified protein detected 7% phosphorylation at Ser1170 (corresponding to Ser1157 in *Sce*CD).

*Cth*CD-CT crystallized in $P3_12_12$ with unit cell constants of a = b = 195.0 Å and c = 189.5 Å and crystals diffracted to a resolution of 7.2 Å. The structure was solved by molecular replacement using a model composed of *Cth*CT and CD_{C2} as search model in Phaser. CD_{C1} and CD_N were placed manually into the resulting maps and the model was refined using rigid-body, domain-wise TLS and B-factor refinement and NCS- and reference model restrained coordinate refinement in Buster to R_{work}/R_{free} of 0.26/0.28. Residues 1033 – 1035, 1134 – 1152, 1214 – 1252, 2188 – 2195 and 2260 – 2297 were not included in the models. Conservation was mapped onto the *Cth*CD-CT crystal structure as for *Cth*CD-CT_{Cter}. MS analysis of purified protein detected 60% phosphorylation at Ser1170 (corresponding to Ser1157 in *Sce*CD).

*Cth*ΔBCCP ACC crystallized in $P6_422$ with unit cell constants of a = b = 462.2 Å and c = 204.6 Å, resolution was limited to 8.4 Å. Structure determination and refinement was performed as for *Cth*CD-CT. Final R_{work}/R_{free} were 0.31/0.34. Although substantial difference density is observed, in accordance with the lack of the BT and BC domains in the model, no defined positioning of these domains could be derived. Additionally, residues 1033 – 1035, 1134 – 1152, 1214 – 1252, 2188 – 2195 and 2260 – 2297 were not included in the models. MS analysis of purified protein detected 70% phosphorylation at Ser1170 (corresponding to Ser1157 in *Sce*CD).

2.5.4 Small-angle X-ray Scattering

Proteins were thawed on ice and dialyzed overnight against 30 mM MOPS pH 7, 200 mM ammonium sulfate, 5% glycerol and 10 mM dithiothreitol. Raw scattering data were measured at SAXS beamline B21 at Diamond Light Source. The samples were measured at concentrations of 2.5, 5 and 10 mg ml^{-1}. Data were processed using the ATSAS package [215] according to standard procedures [216,217]. A slight increase in scattering in the very low-resolution range was observed with increasing protein concentrations, which may be due to interparticle attraction or minor aggregation. Scattering intensities were thus extrapolated to zero concentration using point-wise extrapolation implemented in Primus [218]. Direct comparison of raw scattering curves demonstrates the similarity of *Cth*ACC and *Cth*ΔBCCP, and the derived values such as Rg and Porod Volume match within expected error margins. Molecular mass estimations based on the SAXS-MOW method [219] derive values of 534.7 kDa and 534.0 kDa for *Cth*ACC and *Cth*ΔBCCP, respectively. The relative discrepancies to the theoretical weights of 516.8 kDa (*Cth*ACC) and 503.0 kDa (*Cth*ΔBCCP) are 3.5% and 6.2%, which is in a typical range for this method [219].

2.5.5 Electron Microscopy

Full-length *Cth*ACC was diluted to 0.01 mg ml^{-1} in 30 mM MOPS pH 7.0, 200 mM ammonium sulphate, 5% glycerol, and 10 mM dithiothreitol. Protein sample was adsorbed to a 200 µm copper grid and stained with 2% uranyl acetate. Grids of *Cth*ACC were imaged on a CM-200 microscope (Philips) equipped with a TVIPS F416 4k CMOS camera (Tietz Video and Image Processing Systems). The voltage used was 200 kV; and a magnification of 50000 x results in a pixel size of 2.14 Å. Initial image processing and particle picking was carried out using Xmipp [220,221]. 22309 particles were picked semi-automatically from 236 micrographs with a box size of 300x300 pixels. After extraction, particles with a z-score of >3 were discarded and 22257 particles were aligned and classified into 48 2D class averages using maximum-likelihood target function in Fourier space (MLF2D). After 72 iterations, 4226 additional particles were discarded and the remaining 18031 particles were re-aligned and classified into 36 classes using MLF2D with a high-resolution cutoff of 30 Å. After 44 iterations the alignment converged and class averages were extracted.

2.5.6 *In vitro* Biotinylation and Activity Assay

To ensure full functionality, *Sce*ACC was biotinylated *in vitro* using the *E. coli* biotin ligase BirA. The reaction mixture contained 10 µM ACC, 3.7 µM BirA, 50 mM Tris-HCl, pH 8, 5.5 mM $MgCl_2$, 0.5 mM biotin, 60 mM NaCl, 3 mM ATP, 10% glycerol and the reaction was allowed to proceed for 7 h at 30 °C.

The catalytic activity of phosphorylated and dephosphorylated *Sce*ACC was measured following the incorporation of radioactive ^{14}C into acid-stable non-volatile material [193]. Dephosphorylated ACC was prepared by over-night treatment with λ protein phosphatase (New England Biolabs) of partially purified ACC before the final gel filtration step. The removal of the phosphoryl group from Ser1157 was confirmed by MS. The reaction mixture contained 0.5 µg recombinant ACC in 100 mM potassium phosphate, pH 8, 3 mM ATP, 5 mM $MgCl_2$, 50 mM $NaH^{14}CO_3$ (specific activity 7.4 MBq $mmol^{-1}$) and 1 mM acetyl-CoA in a total reation volume of 100 µl. The reaction mixture was incubated for 15 min at 30 °C, stopped by addition of 200 µl 6 M HCl and subsequently evaporated to dryness at 85 °C. The non-volatile residue was redissolved in 100 µl of water, 1 ml Ultima Gold™ XR scintillation medium (Perkin Elmer) was added and the ^{14}C radioactivity was measured in a Packard Tricarb 2000CA liquid scintillation analyzer. Measurements were carried out in five replicates and catalytic activities were calculated using a standard curve derived from measurements of varying concentrations of $NaH^{14}CO_3$ in reaction buffer.

2.6 Accession codes

Atomic coordinates and structure factors have been deposited in the Protein Data Bank with accession codes 5i6e (*Sce*CD), 5i87 (*Hsa*BT-CD), 5i6f/5i6g (*Cth*CD-$CT_{Cter1/2}$), 5i6h (*Cth*CD-CT), and 5i6i (*Cth*ΔBCCP).

2.7 Author contributions

MH cloned, expressed, purified and crystallized fungal ACC constructs, determined their structure, and carried out SAXS analysis. ES cloned, expressed and crystallized human ACC CD and determined its structure. EM analysis was carried out by ES, MH and AH. SI contributed to structural analysis and figure

preparation. TM designed and supervised work and analyzed crystallographic data, all authors contributed to manuscript preparation.

2.8 Competing Interests

The authors declare no competing interests.

2.9 Acknowledgements

We thank the staff of beamlines X06DA and X06SA at Paul Scherrer Institute, Villigen, Switzerland, for support with crystallographic data collection and the beamline scientists of B21 at Diamond Light Source, Oxfordshire, U.K., for support with SAXS measurements. We thank K. Goldie and H. Stahlberg (C-CINA, Basel, Switzerland) for their support with electron microscopy, the Biophysics facility of the Biozentrum for protein characterization and the Proteomics core facility for mass spectrometry and identification of post-translational modifications. Moritz Hunkeler acknowledges generous support by a Novartis Excellence Fellowship. Anna Hagmann is supported by a “Fellowship for Excellence” of the Biozentrum Basel International PhD Program. This work was supported by Swiss National Science Foundation grants 138262, 159696 and 145023.

2.10 Supplemental Data

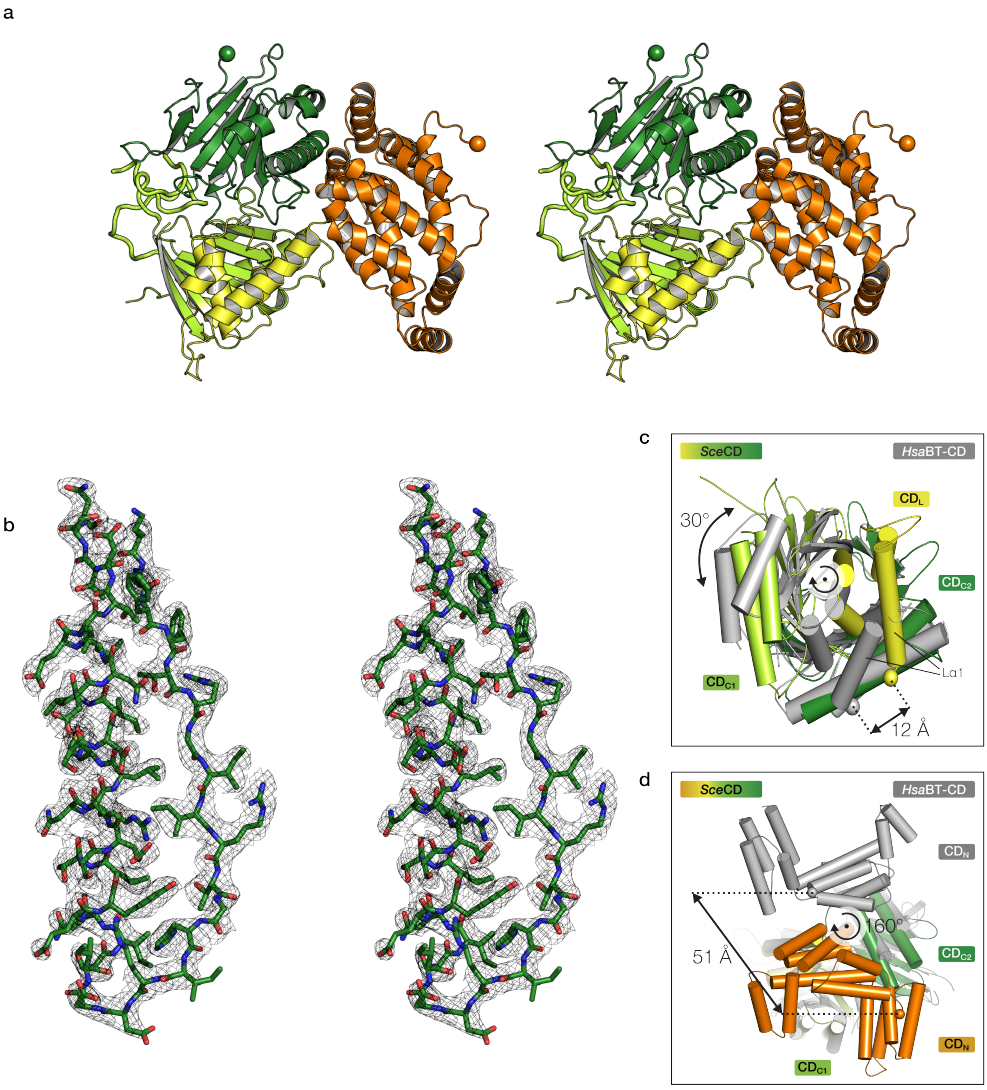
Supplemental Table 2.1 Crystallographic data collection and refinement statistics.

	SceCD	SceCD Thimerosal	SceCD Eu	HsaBT-CD	HsaBT-CD Cd^{2+}	CthCD-CT_{Cter1}	CthCD-CT_{Cter2}	CthCD-CT	CthΔBCCP
Data Collection									
Space Group	$P4_32_12$	$P4_32_12$	$P4_32_12$	$I4_122$	$I4_122$	$P2_12_12_1$	$P2_12_12_1$	$P3_12_12$	$P6_422$
Cell dimensions									
a, b, c (Å)	110.86, 110.86, 131.12	111.22, 111.22, 131.49	108.65, 108.65, 127.36	267.27, 267.27, 210.61	267.67, 267.67, 210.46	97.66, 165.34, 219.23	100.17, 153.45, 249,24	295.02, 295.02, 189.52	462.20, 462.20, 204.64
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 120	90, 90, 120
Resolution[§] (Å)	3.0	3.4	4.0	3.7	4.1	3.6	4.5	7.2	8.4
R_{Merge}*	18.2 (389.6)	20.5 (306.1)	40.6 (327.0)	7.5 (400.9)	15 (730.5)	14.5 (384.5)	27.4 (225.6)	5.6 (302.6)	29.4 (381.7)
CC ½*,[§]	100 (58.3)	99.9 (42.6)	99.9 (48.5)	100 (59.4)	99.8 (73.2)	99.9 (50.9)	99.5 (46.7)	100 (33.3)	99.7 (35)
I / σI*	24.68 (1.46)	7.99 (0.89)	17.92 (1.85)	21.24 (1.07)	16.53 (1.41)	10.61 (0.97)	6.35 (1.00)	18.95 (0.92)	9.05 (0.9)
Completeness*	99.9 (99.9)	99.6 (100)	99.7 (96.8)	99.8 (99.1)	99.8 (99.7)	99.7 (99.9)	99.4 (98.6)	99.6 (100)	99.1 (99.9)
Redundancy*	39.1 (39.8)	12.1 (14.3)	81.6 (65.2)	13.7 (13.7)	20.9 (19.1)	12.7 (13.5)	6.1 (6.5)	9.9 (10.4)	18.5 (18.2)
Refinement									
Resolution	3.0			3.7		3.6	4.5	7.2	8.4
Total Reflections*	662713 (49004)	383712 (64731)	1018078 (58962)	1061661 (78722)	1238529 (84313)	529104 (41183)	142751 (27042)	139139 (10777)	224630 (15976)
Unique Reflections*	16929 (1232)	31675 (4521)	12473 (905)	77629 (5730)	59386 (4406)	41802 (3046)	23225 (4148)	14050 (1040)	12113 (880)
R_{work}/R_{free}	0.20/0.25	-	-	0.38/0.41[†]	-	0.20/0.25	0.26/0.30	0.26/0.28	0.31/0.34
Number of Atoms									
Protein	5543			6843		16330	16330	22582	22582
Waters	76	-	-	-	-	-	-	-	-
Ligand/ion	7	-	-	-	-	-	-	-	-
B-factors									
Protein	129			84		221	268	276	261
Waters	84	-	-	-	-	-	-	-	-
Ligand/ion	93	-	-	-	-	-	-	-	-
R.m.s. deviations									
RMS (angles, °)	1.07	-	-	1.56	-	1.23	1.15	1.14	1.03
RMS (bonds, Å)	0.01	-	-	0.01	-	0.01	0.01	0.01	0.01

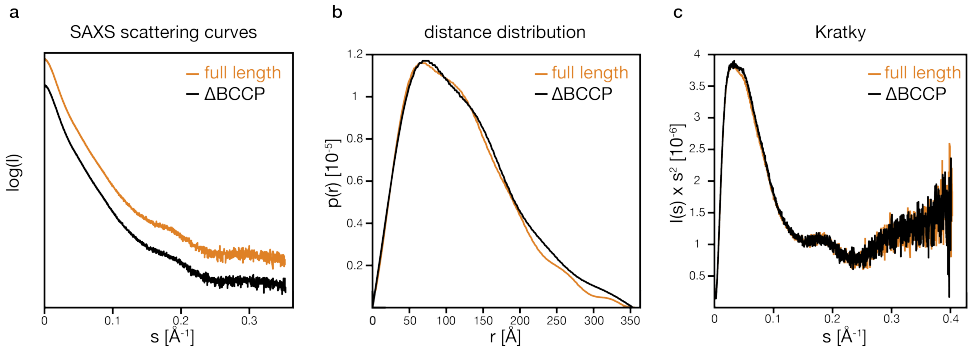
* highest resolution shell is shown in parentheses. [†] Modeled only as poly-alanine, § resolution cut-offs determined based on internal correlation significant at the 0.1% level as calculated by XDS.

Supplemental Table 2.2 Small angle x-ray scattering data collection and processing

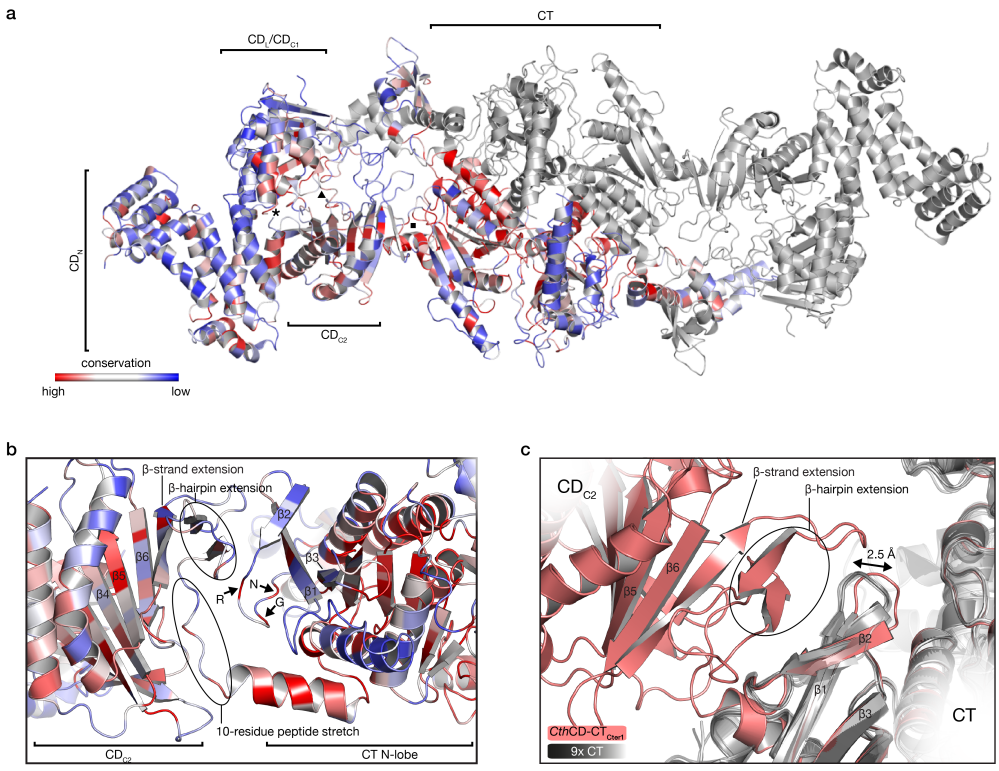
	full length *Cth*ACC	*Cth*ΔBCCP
Instrument	B21 at Diamond Light Source	B21 at Diamond Light Source
Beam size (mm)	1 x 5	1 x 5
Wavelength (Å)	1	1
q range (Å)	0.005 - 0.4	0.005 - 0.4
Detector distance (m)	3.92	3.92
Temperature (K)	293	293
Protein concentration (mg/ml)	2.5, 5, 10	2.5, 5, 10
Scan repeats	30	30
Total exposure time (s)	300	300
Capillary diameter (mm)	1.6	1.6
I(0) (A) [from P(r)]	0.023	0.024
Rg (Å) [from P(r)]	94	97
Rg (Å) (from Guinier)	89 ± 2	92 ± 3
Porod volume estimate (Å)	1176290	1174390
Dmax (Å)	350	350



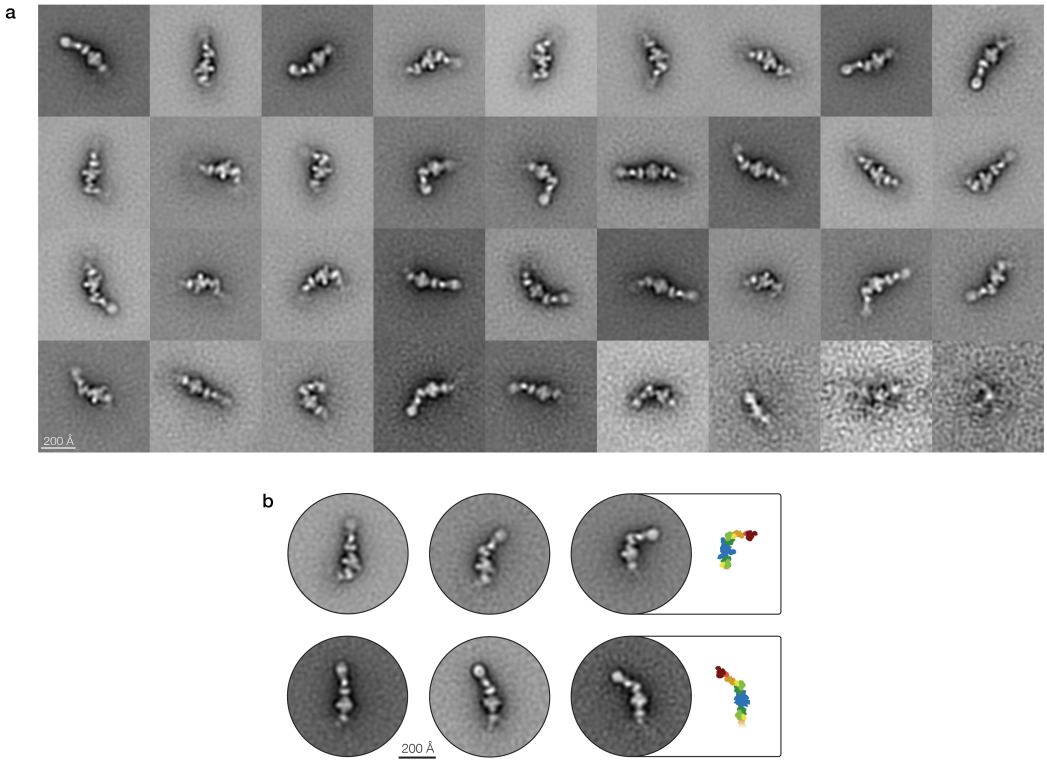
Supplemental Figure 2.1 Structure of *Sce*CD and comparison to *Hsa*BT-CD.
(**a**) Stereo view of *Sce*CD accompanying Fig 1b. (**b**) Stereo view of a representative part of the electron density in CD_{C2} of *Sce*CD. Shown is a 2Fo-Fc map contoured at 1 σ. (**c**) 30° rotation of CD_{C1} in *Sce*CD and *Hsa*BT-CD visualized based on CD_{C2} superposition. *Sce*CD is shown in color and *Hsa*BT-CD is shown in gray. Both CD_N domains and the *Hsa*BT domain were omitted for clarity. The α-helix Lα1 of both structures, the domain rotation and the distance between N-termini (shown as spheres) is indicated. (**d**) A 160° rotation relates CD_N in *Sce*CD to *Hsa*BT-CD, visualized by CD_L/CD_{C1} superposition. *Hsa*BT domain was removed for clarity. Spheres indicate the N-termini and the coloring is as in (c).



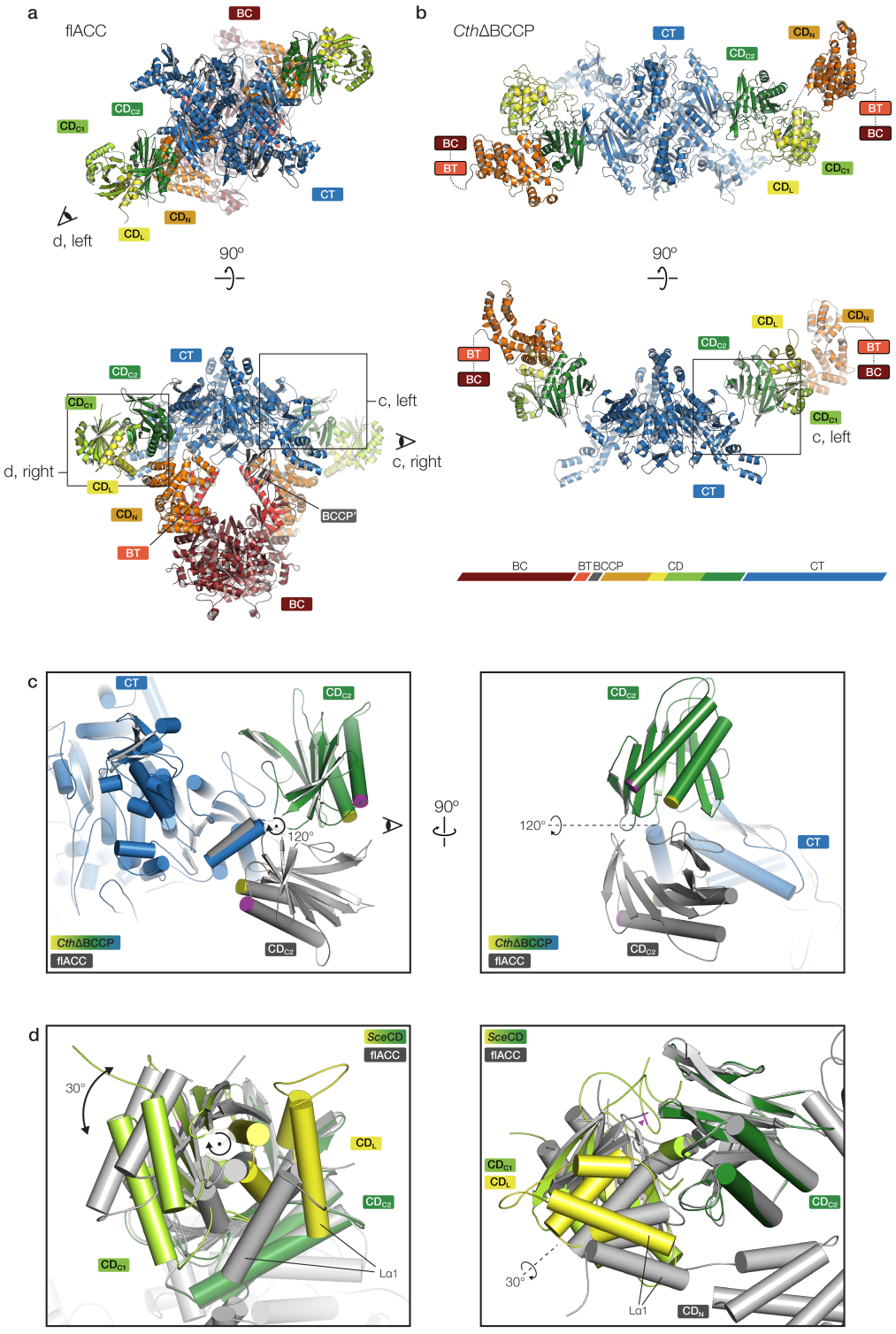
Supplemental Figure 2.2 SAXS analysis of fungal ACC.
Scattering curves (**a**), distance distribution (**b**) and Kratky plots (**c**) of full-length *Cth*ACC (orange) and *Cth*ΔBCCP (black) demonstrating that BCCP-deletion has no impact on the overall solution structure. For clarity, scattering curves are shown with a y-axis offset.



Supplemental Figure 2.3 Conservation analysis of *Cth*ACC.
(**a**) Cartoon representation of *Cth*CD-CT with one protomer colored according to conservation in a gradient from high (red) to low (blue). The second protomer is colored in gray. Contact regions of CD_N and CD_{C1}/CD_L (asterisk), CD_{C1} and CD_{C2} (triangle) and CD_{C2} and CT (square) are surrounded by regions of higher conservation. (**b**) Close-up on the edge-like hinge between CD_{C2} and CT. Highly conserved residues of RxxGxN motif are indicated as well as important strands of CD_{C1} and CT. The second CT protomer is omitted for clarity. (**c**) Superposition of nine isolated CT crystal structures (gray) onto the CT domain of *Cth*CD-CT_{Cter1} (red). The loop preceding the conserved RxxGxN motif is displaced in the CD-CT contact by 2.5 Å compared to isolated crystal structures. Pdb accession codes of structures used for superposition: 1od2, 1uys, 1uyt, 1w2x, 3h0j, 3h0q, 3k8x, 3pgq, 3tv5.

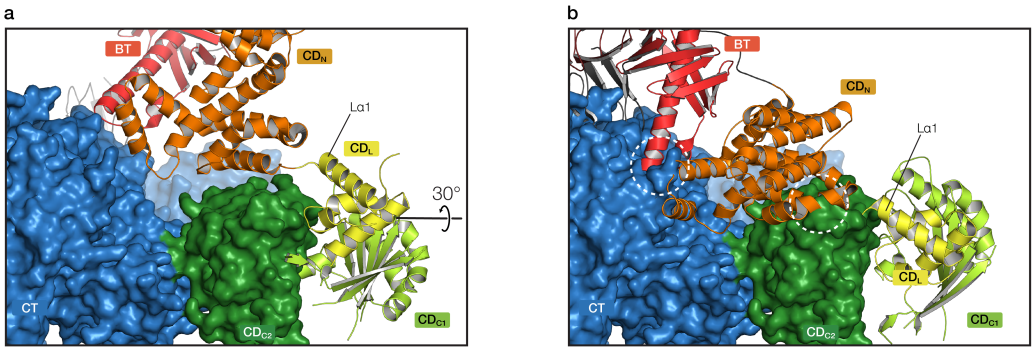


Supplemental Figure 2.4 Negative stain electron microscopy class averages of *Cth*ACC illustrating its flexibility.
(**a**) 18031 particles were 2D-classified into 36 classes using the MLF2D method. Classes are sorted by number of particles per class, with the highest number at the top left and the lowest number at the bottom right. Scale bar 200 Å. (**b**) 2D-class averages from negative stain EM analysis depict the ensemble of solution conformations of the intrinsic flexibility of *Cth*ACC. Scale bar 200 Å.

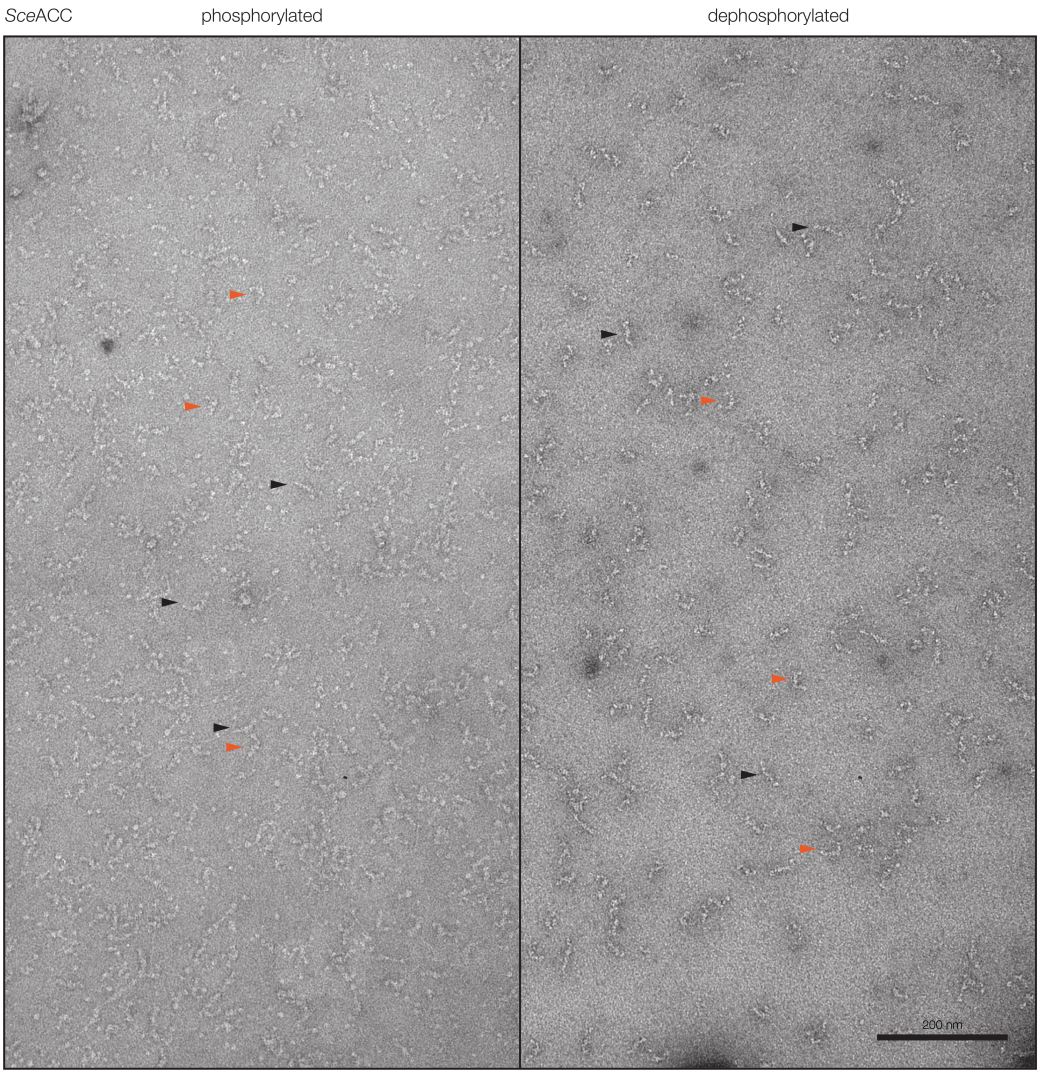


Supplemental Figure 2.5 Conformational variability in fungal ACC.
(**a**) Cartoon representation of triangular flACC (pdb accession code: 5csl). The domains are labeled according to the color scheme in (b). (**b**) Cartoon representation of *Cth*ΔBCCP colored according to the given color scheme. The orientations of (a) and (b) are based on a superposition of the CT domains. Fields of view for subpanels (c) and (d) are indicated with boxes and eyes. (**c**) Left: Detailed view of the CD_{C2}/CT 120° hinge motion viewed along the rotation axis. *Cth*ΔBCCP is colored and flACC is shown in gray for clarity.

Corresponding ends of CD_{C2} helices are colored in yellow and magenta. The rotation axis is indicated and the view for the right subpanel is indicated with an eye. Right: View onto the rotation axis. (**d**) Conformational states of the CD_{C1}-CD_{C2} hinge in the non-phosphorylated flACC and in the phosphorylated *Sce*CD crystal structures when superimposed on CD_{C2} domains. Lα1, the first N-terminal helix of the CD_L domain, is labeled to illustrate the conformational differences. The left panel shows the view along the rotation axis, the right panel corresponds to a view onto the axis. The position of the phosphorylated Ser1157 in the regulatory loop is indicated by a magenta triangle.



Supplemental Figure 2.6 Conformation of phosphorylated CD_{C1}-CD_{C2} hinge is incompatible with BT/CD_N conformation in flACC.
(**a**) Detailed view on flACC (PDB accession code: 5csl) with CT and CD_{C2} domains shown as surface and BT, BCCP, CD_N, CD_L and CD_{C1} domains shown as cartoon. (**b**) flACC model with the CD_{C1}-CD_{C2} hinge conformation as observed in the *Sce*CD crystal structure. Clashes of BT/CD_N with CT and CD_N with CD_{C2}, that are caused by the 30° rotation of CD_{C1} relative to CD_{C2} (rotation axis shown in (a)), are indicated with white circles. For orientation, the first helix of CD_L is labeled in both panels.



Supplemental Figure 2.7 Negative stain electron microscopy of phosphorylated and dephosphorylated *Sce*ACC shows similar degree of conformational flexibility. Micrographs of phosphorylated (left) and dephosphorylated *Sce*ACC (right) are shown side by side. Selected particles representing elongated molecules and U-shaped molecules are indicated with black and orange arrows, respectively. The scale bar is 200 nm.

3 Hybrid Structure of a Bacterial Single-Chain Carboxylase

Based on (manuscript in preparation):

Hybrid structure of a single chain carboxylase from *Deinococcus radiodurans.*

Anna Hagmann, Moritz Hunkeler*, Edward Stuttfeld, Timm Maier*

3.1 Abstract

Biotin-dependent acyl-Coenzyme A (CoA) carboxylases (aCCs) are involved in key steps of anabolic pathways and comprise three distinct functional units; biotin carboxylase (BC), biotin carboxyl carrier protein (BCCP) and carboxyl transferase (CT). YCC multienzymes are a poorly characterized family of prokaryotic aCCs of unidentified substrate specificity, which integrate all functional units into a single polypeptide chain. We employed a hybrid approach to study the dynamic structure of *Deinococcus radiodurans* (Dra) YCC: crystal structures of isolated domains reveal a hexameric CT core with extended substrate binding pocket and a dimeric BC domain. Negative stain electron microscopy provides an approximation for the variable positioning of the BC dimers relative to the CT core. Small-angle X-ray scattering yields quantitative information on the ensemble of Dra YCC structures in solution. Comparison to other carrier protein-dependent multienzymes highlights a characteristic range of large-scale interdomain flexibility in this important class of biosynthetic enzymes.

3.2 Introduction

Biotin-dependent carboxylases are ubiquitous enzymes catalyzing ATP-dependent carboxylation in fatty acid, carbohydrate and amino acid metabolism, as well as in urea utilization and microbial polyketide biosynthesis [42,43,222-224]. Based on their substrate specificity, biotin-dependent carboxylases are classified into three families: urea carboxylases (UCs), pyruvate carboxylases (PCs) and the general family of acyl-CoA carboxylases (aCCs) [57,62,225]. The aCC family includes acetyl-CoA carboxylase (ACC), which catalyzes the key committed step in fatty acid biosynthesis, propionyl-CoA carboxylase (PCC) and 3-methylcrotonyl-CoA carboxylase (MCC) [42,226-229]. Despite their diverse substrate specificities, all biotin-dependent carboxylases share a common enzymatic mechanism and domain organization. Carboxylation is carried out in two half reactions and involves three functional components. First, a biotin carboxylase (BC) catalyzes the ATP-dependent carboxylation of a biotin cofactor, which is covalently linked to a conserved lysine of the biotin carboxyl carrier protein (BCCP). Then, BCCP translocates to the carboxyl transferase (CT), where the carboxyl group of the carboxybiotin intermediate is transferred to the respective substrate [10,230]. In this reaction scheme, the BC and BCCP are conserved components, while the CT varies in active site structure depending on the substrate. The structural organization differs vastly between various aCCs, despite their conserved reaction logic. aCCs either occur as multi-subunit enzymes, where the enzymatic functions are provided by distinct protein subunits, or as single-chain multienzymes, which integrate all enzymatic domains into one polypeptide chain (Figure 3.1A). The group of multi-subunit aCC comprises most prokaryotic forms, which consist of two, three or four subunits as exemplified by *Streptomyces coelicolor* PCC, *Metallosphaera sedula* ACC or *Escherichia coli* ACC,

respectively, as well as the eukaryotic PCC and MCC. Single-chain aCCs are represented by eukaryotic ACCs, which comprise a large, non-catalytic central domain in addition to the canonical enzymatic domains. Strikingly different oligomeric assemblies are observed for aCCs: Bacterial PCC and MCC are hetero-dodecamers with $(BC\text{-}BCCP)_6(CT)_6$ stoichiometry [228,229]. Bacterial ACCs have a $(BC)_2(BCCP)_4(CT_\alpha CT_\beta)_2$ stoichiometry [34] and the archaeal ACC holoenzyme forms a $(BC)_4(BCCP)_4(CT)_4$ hetero-dodecamer [83]. Eukaryotic ACCs function as dimers or higher oligomeric filaments [85,194,231,232].

Recently, in bacteria, a group of single-chain aCCs of yet unknown substrate specificity was identified. These enzymes, denoted as YCCs, encompass all functional domains in a single polypeptide chain of ~1200 amino acids. They share the domain order with eukaryotic ACCs, but lack the large central domain (Figure 3.1A). YCCs occur in a diverse set of unrelated bacterial species, including *Pseudomonas aeruginosa, Cupriavidus metallidurans and Deinococcus radiodurans*. For most YCCs, neither the substrates nor their physiological function or structure have been identified. The only exceptions are *Mycobacterium avium* LCC (Map LCC) and *Rhodopseudomonas palustris* LCC (Rpa LCC), for which a crystal structure and an initial substrate characterization, respectively, have been obtained recently [233]. Based on high-resolution crystal structures of the CT and BC subunits and solution structural data, we report a hybrid model of the *D. radiodurans* YCC (Dra YCC), which provides insights into active site architectures, domain interactions and the dynamic organization of a complex YCC multienzyme.

3.3 Results

Dra YCC was overexpressed in Sf21 insect cells and purified in biotinylated form with a yield of ~30 mg per liter of culture. It forms a hexamer of ~680 kDa in solution, as confirmed by size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) (Supplemental Figure 3.1A). Size exclusion chromatography and SDS-PAGE analysis of the protein confirmed homogeneity and purity of the sample (Supplemental Figure 3.1B and C). We employed a hybrid approach of determining high-resolution crystal structures of the individual enzymatic domains in combination with analyzing the overall assembly structure and flexibility in solution by small-angle X-ray scattering (SAXS) and electron microscopy (EM).

3.3.1 Crystal structure of a ring-shaped hexameric CT domain assembly

We overexpressed the CT domain of Dra YCC (amino acids 556 – 1091) in Sf21 cells. The protein forms a hexamer of ~330 kDa in solution based on SEC-MALS analysis (Supplemental Figure 3.1D). The CT domain was crystallized in space group $P2_12_12_1$ and its structure was determined by molecular replacement using the transcarboxylase 12S subunit [82] as search model. The final model comprising residues 573 – 1091 was refined to R_{work}/R_{free} at 20.2%/24.1% at 2.4 Å resolution (Supplemental Table 3.1). The CT domain consists of two subdomains, the N-lobe (residues 554 - 832) and the C-lobe (residues 833 – 1091). Consistent with the oligomeric state in solution, the CT domain forms a hexameric ring with D3 symmetry, consisting of a trimer of dimers with an outer diameter of ~140 Å and a height of ~67 Å (Figure 3.1B and C). Dimerization of CT is mediated by head-to-tail interaction of two protomers with an interface area of ~3900 $Å^2$. Both

subdomains contribute to trimerization of CT dimers via an interface area of ~2800 Å^2 per dimer.

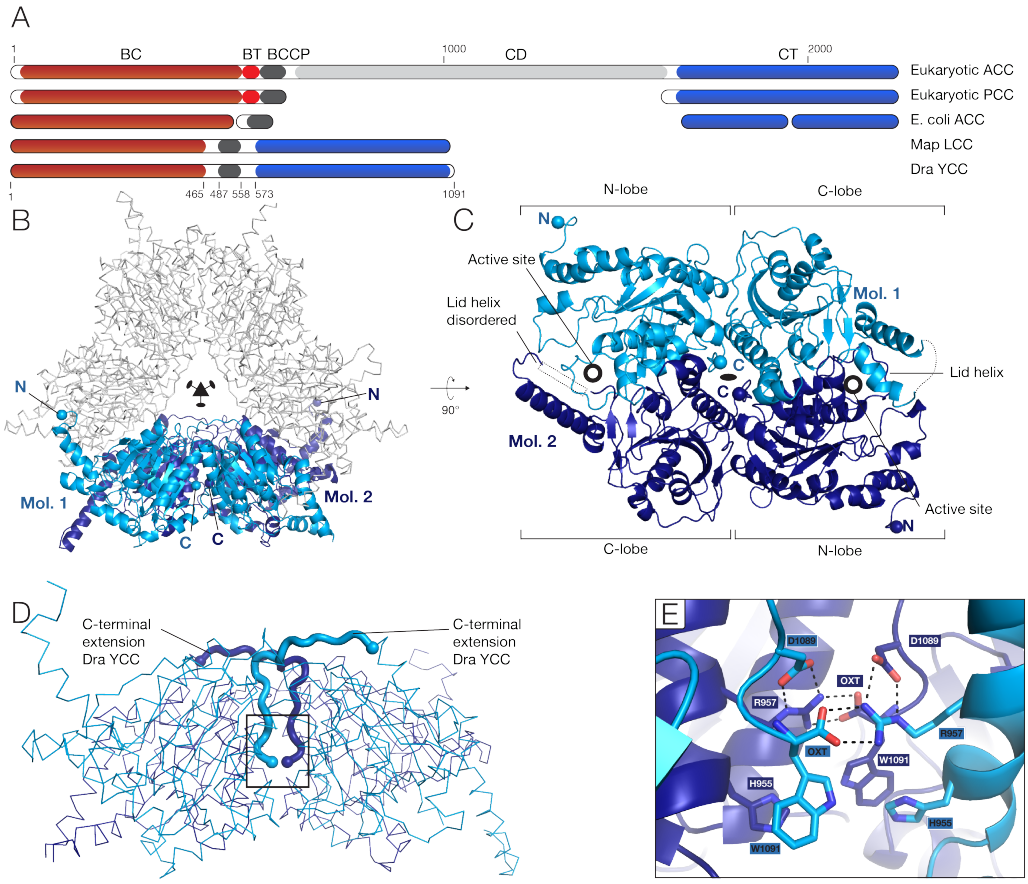


Figure 3.1 Dra YCC domain organization and hexameric CT domain structure.
A. Schematic domain organization of eukaryotic ACC, PCC, *E. coli* ACC, Map LCC, and Dra YCC. Domain boundaries of Dra YCC are indicated. BC is shown in ruby, BCCP in dark gray, CT in blue, BT in red and CD domains in light gray. **B.** Top view of the hexameric CT domain organized as trimer of dimers. One dimer is depicted in light and dark blue, the remaining structure in gray, the D3 symmetry is indicated. N- and C-termini of the colored dimer are indicated with spheres. **C.** Side view of one CT dimer, as indicated in B. The dimer exhibits an antiparallel arrangement with a central 2-fold axis and active sites located at its distal interfaces. Disordered regions are indicated in dotted lines and N- and C-lobe are indicated by brackets. **D.** Dra YCC CT dimer with stabilizing C-terminal extensions highlighted as bold loops. Beginning of extensions and C-terminus are indicated by spheres. View is the same as in C, only one dimer is shown for clarity. **E.** Interactions of the conserved C-terminal DXW motif. View as indicated by a rectangle in D, top protomer is clipped for clarity. Salt bridges are shown as black dashed lines. See also Supplemental Figure 3.1 and 3.2.

3.3.2 A conserved C-terminal peptide extends the CT dimer interface

A comparison of the Dra YCC and Map LCC CT structures [233] shows conservation of oligomeric interaction and overall assembly shape. However, a

major difference is observed in the inter-subunit dimer interface: Dra YCC features a 15 amino acid C-terminal extension (Figure 3.1D), which enlarges the dimer interface by ~700 Å^2 relative to Map LCC. The C-terminal extension protrudes into the dimer interface from the inside of the hexameric ring, while Map LCC already terminates at the end of a conserved helix (Supplemental Figure 3.2A). Based on sequence alignment, a C-terminal extension of ~10 to 20 amino acids with a highly conserved aspartate-X-tryptophan (DXW) motif at the very C-terminus is present in 20 out of 29 YCC sequences (Supplemental Figure 3.2B). The C-terminal DXW motif mediates Dra YCC intersubunit interactions: the motif's aspartate 1089 and the C-terminal carboxyl group are both engaged in bidentate interactions with the conserved arginine 957 of the neighboring protomer (Figure 3.1E). Arginine 957 is also found in Map LCC, which lacks a C-terminal extension, and there forms a salt bridge to aspartate 919 in a neighboring helix from the same subunit. The tryptophan of the DXW motif, stacks with histidine 955, which is conserved in all but one of the examined YCCs.

C-terminal core fold extensions of similar length and with distinct conserved C-terminal motifs are also present in the two other families of structurally characterized hexameric CT domains, namely PCCs and MCCs (Supplemental Figure 3.2C and D) [229,234]. In PCC, these extensions extend into the same region of the hexameric CT assembly and terminate at the equivalent position of the Dra YCC C-terminal extension. However, in PCC the extensions are not crossed over (Supplemental Figure 3.2C). In MCC, which is characterized by a domain-swapped organization of the CT N- and C-lobe relative to YCC and PCC, the C-terminal extensions adopt a completely different topology and protrude into the interface responsible for dimer trimerization (Supplemental Figure 3.2D) [229]. As in YCC and PCC, the extension only spans about 2/3 of the interface. Overall, C-

terminal core fold extensions are observed in all structurally characterized hexameric CT domains, except for the Map LCC-like subclass of YCCs. In these systems, the C-terminal extensions are located in intersubunit or interdomain interfaces. However, their topology and the interactions of their distinct conserved C-terminal motifs are strikingly divergent.

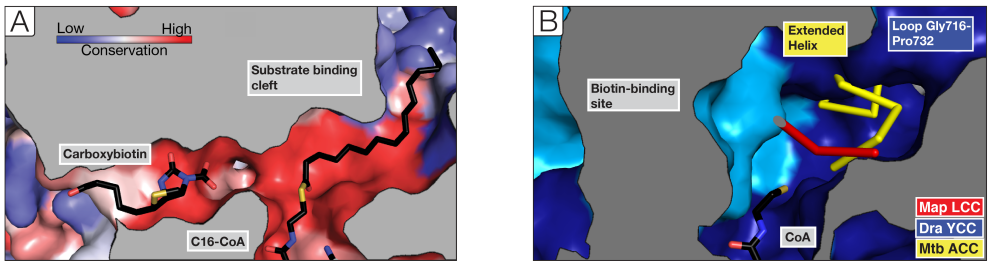


Figure 3.2 Active site of Dra YCC CT.
A. The extended substrate binding cleft of the Dra YCC CT domain is lined with conserved residues. Color gradient from blue to red indicates increasing conservation; for visualization, a CoA moiety is modeled according to its positioning in the related MCC holoenzyme (PDB ID: 3U9R [228]) and the carboxybiotin and palmitoyl are docked into active site pockets. **B.** Overlay of the CT active sites of *M. tuberculosis* (Mtb) ACC, Map LCC and Dra YCC. Dra YCC CT is shown surface representation (light and dark blue for two protomers), loops occluding the binding site tunnel of Mtb ACC and Map LCC are shown in yellow and red, respectively [233,235]. See also Supplemental Figure 3.3.

3.3.3 The active site cleft in Dra YCC CT is extended

The CT active site is formed at the dimer interface between the N-lobe of one protomer and the C-lobe of its dimer partner (Figure 3.1C). It consists of a biotin- and a substrate specific binding pocket. The substrate binding pocket of Dra YCC is an elongated and narrow cleft of ~20 Å length (Figure 3.2A) and a volume of ~2000 Å^3, which is lined by conserved residues and sufficient in length for enclosing a C_{16} acyl chain with possible space for accommodation of longer or branched substrates. The binding groove is much more extended than in ACCs with their small acetyl substrate, *e.g. M. tuberculosis* ACC CT [235] (Figure 3.2B and S3A). The active site region of Dra YCC is rather conserved at the sequence level to Rpa LCC, which is active towards C_2 to C_{16} substrates, but with a preference for long-chain acyl chain substrates [233]. However, in the crystal structure of the

related Map LCC, the substrate binding cleft is occluded by a loop formed by residues 715 – 732 (Figure 3.2B and S3B), such that a requirement for conformational changes during substrate accommodation has been suggested [233]. The corresponding loop glycine 716 – proline 732 in Dra YCC adopts a different conformation, which results in an opening of the substrate binding groove. Dra YCC may thus represent a substrate-binding competent form of the active site. In Map LCC, a helix spanning residues 781-787 occludes the CoA binding region of the active site [233]. Although the positioning of the corresponding helix (residues 781 – 788) is well conserved in the Dra YCC CT structure, already a slight shift of the CoA moiety relative to its position observed in the crystal structure of MCC, would be sufficient for accommodating CoA in an equivalent pocket also in Dra YCC CT. The respective regions are not constrained by crystal contacts in either the Dra YCC CT or Map LCC CT crystal structures. A helical hairpin flap (amino acids 1005 – 1046) with increased flexibility is positioned above the active site of Dra YCC. In the structure of the homologous *P. aeruginosa* MCC CT domain in complex with CoA, the corresponding flap acts as a lid helix by closing down onto the bound CoA moiety [229], suggesting a similar role in CoA and substrate accommodation also in Dra YCC (Figure 3.1C).

3.3.4 Limited proteolysis enables structure determination of dimeric BC

Initial expression screening failed to yield a solubly expressed construct of the Dra YCC BC domain. However, a stable BC domain fragment was obtained by limited proteolysis of full-length Dra YCC. Initial crystals of the fragment were affected by merohedral twinning with a twin fraction close to 50%, which precluded structure determination. Through extensive screening, a crystal with a twin fraction of ~30% in space group $P6_5$ with two molecules in the asymmetric unit was obtained and

the Dra YCC BC domain structure solved using molecular replacement with the BC domain of PC [236]. The final model consists of two almost identical monomers (RMSD of 0.14 Å) (Supplemental Figure 3.4A) including residues 1 - 464 and was refined to R_{work}/R_{free} of 14.9%/18.4% at 1.7 Å resolution using twin refinement in REFMAC5 [237].

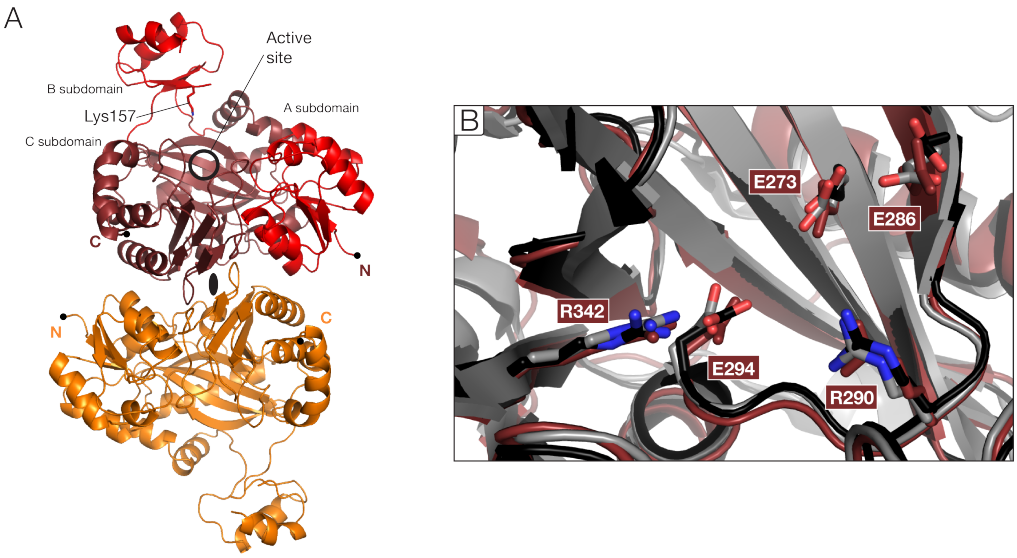


Figure 3.3 Dimerization and active site structure of the BC domain.
A. Cartoon representation of the dimeric BC domain structure, one protomer is shown in red, the other in orange, the two-fold symmetry dimer axis is indicated. The three subdomains are shown in shades of red in one protomer. The lid in the B subdomain is in an open conformation and partially disordered. The active site location is marked in one protomer. **B.** The active site structure of the Dra YCC BC domain superimposed onto *E. coli* ACC (PDB ID: 3G8C [238]) in gray and *R. pomeroyi* PCC (PDB ID: 3N6R [228]) in black. Catalytic residues are labeled; residue numbering is according to Dra YCC BC. Conservation of the BC active site indicates a conserved reaction mechanism of Dra YCC. See also Supplemental Figure 3.4.

The BC domain consists of the rigid A and C subdomains and the small interspersed B subdomain (Figure 3A) [239]. The B subdomain is flexibly tethered by two irregular linkers and exhibits increased disorder in the crystal (Supplemental Figure 3.4B). Based on homology to the BC domains of *E. coli* or *Haemophilus influenzae* ACC, this B subdomain may undergo a hinge-bending motion during substrate binding to act as a lid to the BC active site [240,241]. Particularly the highly conserved glycine-rich loop (lysine 157 – methionine 167) in the B subdomain

might be involved in direct ligand contacts, with lysine 157 presumably interacting with one of the α-phosphoryl oxygens of MgATP (Figure 3A).

The BC active site is located at the interface of the A and C subdomains. Active site residues are well conserved between YCC, ACC [239] and PCC BC [228] (Figure 3B). Structural homology suggests that residues glutamic acid 273 and glutamic acid 286 in Dra YCC BC coordinate magnesium ions, which are presumably involved in binding of the ATP substrate. Arginine 290, glutamic acid 294 and arginine 342 are directly implicated in binding of biotin and ATP [228,239,241]. Carboxybiotin-interacting residues have been identified in the crystal structure of the *H. influenzae* ACC BC [241] and are structurally conserved in Dra YCC (lysine 235, arginine 290 and arginine 342) (Figure 3B), in agreement with a generally conserved BC mechanism.

The BC domains form dimers with an interface area of ~900 $Å^2$ and a maximum extent of 100 Å, analogous to the *E. coli* and *H. influenzae* ACC BC as well as the BC domain of the Map LCC [233,238,241]. Contrastingly, the BC domains of eukaryotic ACC are monomers in isolation, but they may form a permanent or transient dimer in context of the full-length ACC multienzyme [194,232,242].

3.3.5 Negative stain electron microscopy reveals variable BC domain positioning

To gain insight into the overall Dra YCC assembly, we employed negative stain electron microscopy (EM). Dra YCC particles are characterized by a central triangular shape, which corresponds to the hexameric CT domain (Figure 3.4A). 9477 particles were picked and used for the generation of 24 2D class averages (Supplemental Figure 3.5). 16 of these 24 classes represent top views of the CT hexamer with laterally positioned dimeric BC domains (Figure 3.4B), only one

class provides a side view along a twofold symmetry axis, the remaining classes are projections showing the CT domain tilted to ~45°. Projections of the CT crystal structure correspond well to the respective class averages (Figure 3.4C).

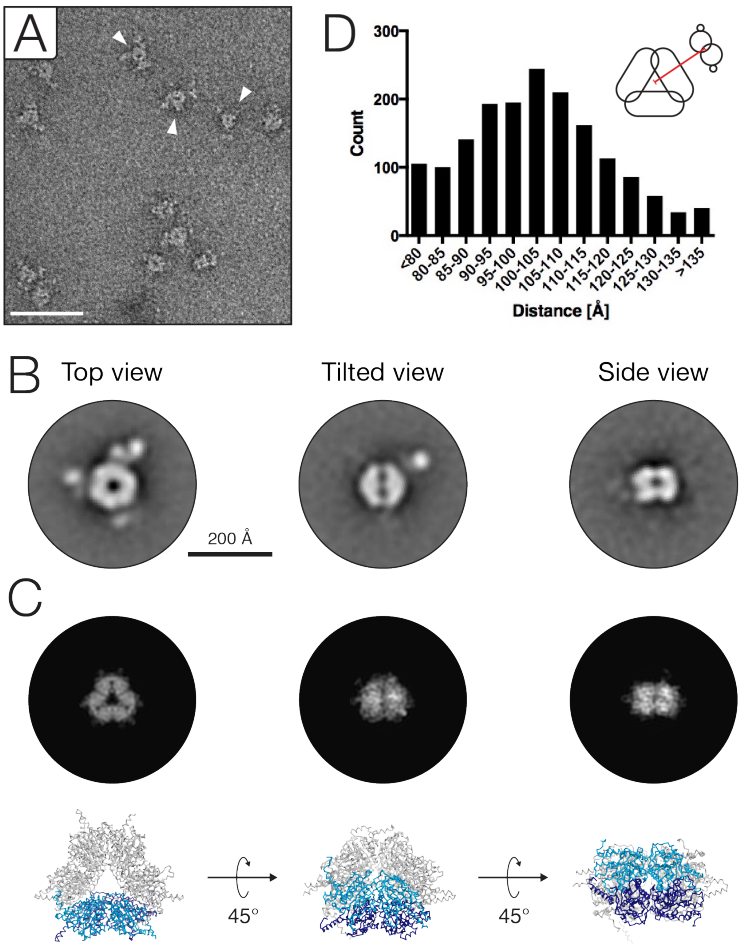


Figure 3.4 Negative stain EM analysis of BC domain mobility.
A. Cropped raw negative stain EM micrograph. Three Dra YCC particles are indicated with white arrows. Scale bar is 500 Å. **B.** Representative negative stain EM 2D class averages of Dra YCC in top view (left), tilted view (middle) and side view (right). In the top view, the hexameric CT ring is clearly visible; the BC dimers are positioned laterally to the ring. **C.** Projections of the Dra YCC CT crystal structure, filtered to 20 Å resolution, as well as images of the structure, corresponding to the class averages in B, are shown. **D.** Histogram of measured BC-CT distance in distance classes of 5 Å increments demonstrating the variable positioning of the BC domains relative to the CT domain hexamer. See also Supplemental Figure 3.5.

Whilst the CT hexamer is preserved in all classes, significant variability is observed in positioning of the BC domains. To further analyze BC positioning, we measured the pixel distances between the center of the CT hexamer and the respective BC dimers in particles of top view classes as an approximation of CT-

BC distance. ~1600 distance measurements were sorted into classes of 5 Å increments (Figure 3.4D). The resulting distance histogram demonstrates that BC dimer positioning is highly variable and covers the total range from less than 80 Å, which corresponds to the BC dimer resting on the CT ring as observed in the Map LCC crystal structure [233], to more than 135 Å from the center of the CT ring, which roughly corresponds to maximally extended interdomain linker conformation. Only about 12% of all BC domains were detected in close proximity with the CT domains at distances to the CT center of < 85 Å. The distance distribution has a maximum at an intermediate distance of 105 Å. However, due to measurement in projection images, this value probably still underestimates the real BC-CT distance.

3.3.6 Small-angle X-ray scattering provides an ensemble model of Dra YCC

Negative stain EM indicates a considerable flexibility of the BC domains relative to the CT ring in Dra YCC (Supplemental Figure 3.5). To assess this flexibility under physiological conditions in solution, we subjected Dra YCC to SAXS measurements (Supplemental Table 3.2). Interatomic distance distributions derived from raw scattering data yielded D_{max}=256 Å, while the corresponding value for a Dra YCC model with the BC domains in close proximity of the CT hexamer would only be ~200 Å. A theoretical scattering curve calculated for a symmetrical Dra YCC assembled based on the crystal structure of Map LCC does not provide a good fit to the experimental data (χ^2=3.7, Figure 3.5A). The solution scattering ensemble data thus reinforce the results of the single-particle EM analysis, which suggested a highly variable BC positioning. It was therefore unlikely, that a single, symmetrical Dra YCC model would be sufficient to fit the SAXS data. Therefore, we applied the ensemble optimization method (EOM) for

SAXS data interpretation [243,244]: First, a pool of 10 000 random models was generated by keeping the CT domain fixed while BC dimers were allowed to move freely, only restrained by the approximate linker length. This was achieved by assembling a hexameric Dra YCC model from rigid CT and BC domain dimers linked by a single artificial linker. This linker represents the N- and C-terminal BCCP linkers and was allowed to adopt random, native-like conformations. Then, a minimal ensemble of models with optimal fit to the experimental data was selected [243,245]. This ensemble comprised three models (Figure 3.5B) and yields an improved quality of fit to experimental data as documented by a χ^2 value of 1.45. The D_{max} of 255 Å for the calculated combined scattering curve corresponds well to experimental data. Each of the three models individually exhibits asymmetric BC positioning with variable distances and angles between BC dimers and the CT hexamer.

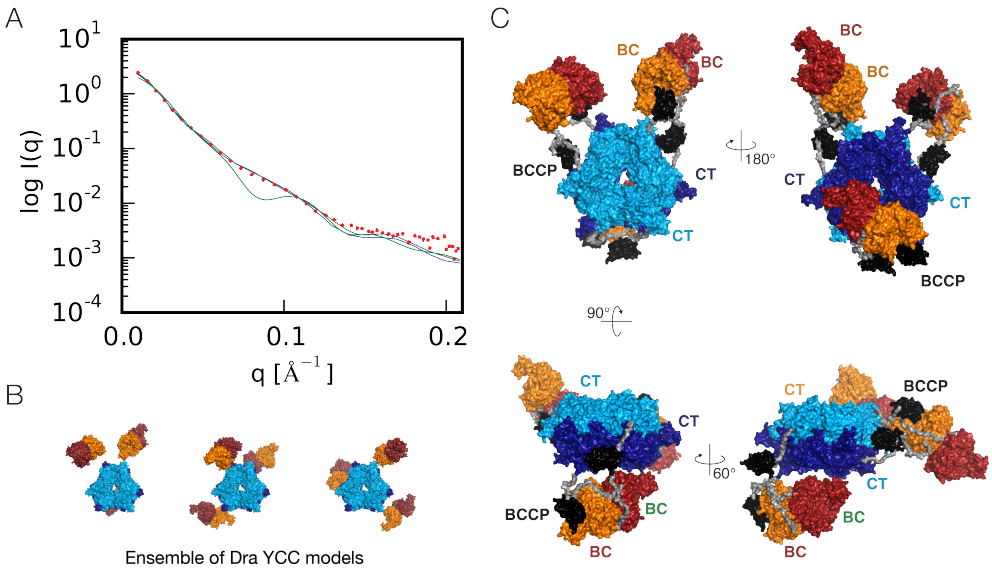


Figure 3.5 SAXS analysis of Dra YCC.
A. Comparison of an experimental scattering curve (red) with calculated scattering curves for a symmetrical Dra YCC model based on Map LCC (green), the EOM-derived three model Dra YCC ensemble (blue) and of a single reconstituted asymmetric Dra YCC model (black, as shown in C). **B.** Ensemble of Dra YCC models generated by EOM. **C.** Reconstructed Dra YCC model featuring asymmetric placement of BC domains. The BCCP domains were modeled according to the crystal structure of the MCC BCCP (PDB ID: 3U9S [229]). The BC dimer is colored in orange and red, the CT monomers in light and dark blue. The BCCP domains are colored in black and the linkers in gray. See also Supplemental Table 3.2.

The EOM derived models only contain an artificial single-stranded linker between a BC dimer and the CT. To obtain a most realistic representation of Dra YCC, we reconstructed a complete Dra YCC model with all linkers and BCCP based on the asymmetric model with largest weight in EOM (Figure 3.5C): The linkers were modeled in stereochemically plausible conformations incorporating the native domain swap suggested for Map LCC (Figure 3.5C) [233], although also non-swapped linker arrangements would be plausible. The BCCP domain was modeled according to the crystal structure of MCC BCCP (PDB ID: 3U9S [229]). Scattering curves calculated from the resulting model provided a good fit to experimental data at a χ^2 value of 0.65 (Figure 3.5A). In the reconstituted Dra YCC full-length model, two of the three BC dimers are in distal positions outside the plane defined by the CT hexamer, while one lies in the same plane. One of the dimers is located inside the ring diameter, outside of the ring plane, while the other two are located at center-to-center distances of 103 Å and 125 Å to the CT hexamer. Altogether, the EOM-based SAXS model is highly consistent with individual particles and class averages visualized in EM and demonstrates that a single symmetric model is not sufficient to provide an adequate representation of the solution conformation of Dra YCC.

3.4 Discussion

The structural analysis of Dra YCC establishes a classification of bacterial single-chain carboxylases into two subfamilies based on the presence of a C-terminal extension that contributes to inter-subunit interactions. Although the cognate physiological substrate of Dra YCC remains unknown, the CT extended active site cleft hints towards a long-chain acyl substrate. This hypothesis is further substantiated by the specific genomic organization of the Dra YCC coding gene

Dra_A0310 into one operon with gene Dra_A309, annotated as a long-chain fatty acid CoA ligase. When examining the genetic context for 21 identified YCCs, ten of them also contain a related ligase in their operon, amongst them *M. avium*, of which the Map LCC structure is known [233]. These organisms are not phylogenetically related, however, the genetic synteny of the YCC and the ligase points towards a common specialized, yet unidentified substrate and metabolic pathway. Genetic and functional linking between a biotin-dependent carboxylase (AccD4) and a protein from the long-chain fatty acid CoA ligase family (FadD32) is for example also observed in operon mtu-Rv3801c of *Mycobacterium tuberculosis*. There, AccD4, a putative carboxyltransferase, carboxylates acyl-CoA substrates in cooperation with AccD3, a biotin carboxylase, not localized on the operon. FadD32 is required for the activation of a meromycolic acid [246]. The third enzyme on the operon, Pks13, then catalyzes the condensation of the meromycolic acid and the carboxylated long-chain fatty acid into a mycolic acid membrane lipid. Speculatively, the YCC or LCC in cooperation with the ligase could be part of a system for activation of long-chain fatty-acid precursors for condensation in the course of the biosynthesis of complex membrane lipids.

By combining negative stain EM to analyze individual particles and SAXS as an ensemble technique, we have obtained a consistent quantification of the remarkable mobility of the Dra YCC BC domain. Large-scale conformational dynamics have been observed in other carrier protein-based enzymatic systems, such as animal fatty acid synthase (FAS), microbial modular polyketide synthases (PKS), and the pyruvate dehydrogenase complex (PDC) (Figure 3.6) [247-249]. Still, substrate transfer in these even more complex systems has commonly been discussed only in terms of the mobility of a flexibly tethered carrier protein. The recognition of the conformational dynamics of BC in single-chain carboxylases

with only two types of enzymatic subunits suggests a more general role of overall assembly dynamics for carrier protein-mediated substrate transfer.

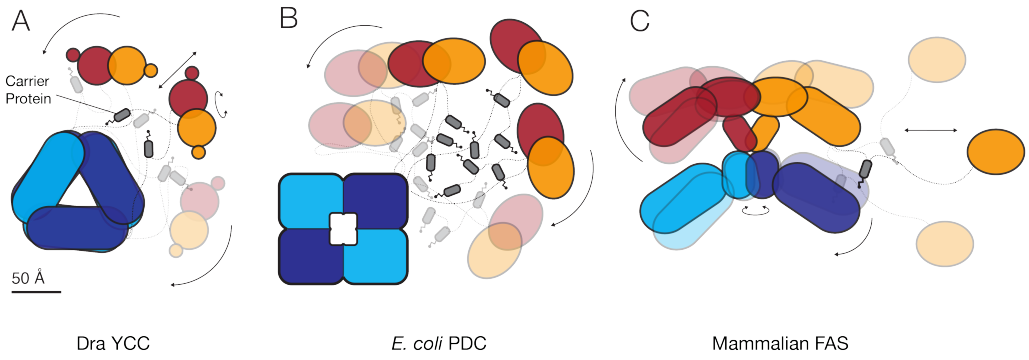


Figure 3.6 Dynamics of carrier protein-dependent multienzyme families.
Rigid regions are shown in light and dark blue, while variable regions are shown in orange and red. Transparent domains show the approximate range of motion of the domains. Arrows represent example modes of domain motion. **A.** Schematic representation of Dra YCC showing only a single instance of BCCP and BC dimer. **B.** Schematic representation of *E. coli* PDC. Only three E3 subunits are shown for clarity. Distances according to Murphy *et al* [20]. **C.** Schematic representation of mammalian FAS with acyl carrier protein (ACP), thioesterase (TE) domain and the linker tethering them to the enzyme [250]. Only one ACP and TE are shown for clarity.

The mode of carrier linkage of the Dra YCC BCCP resembles those of other carrier protein-dependent enzymes: The BCCP-CT linker has a length of 21 amino acids and no preferential sequence composition, it resembles linkers in animal FAS and PKS, with lengths of 12 amino acids in FAS and 15 to over 40 residues in PKS [251]. The BC-BCCP linker is 22 amino acids long and rich in alanine and proline residues. Similar linker compositions are observed in PDC [8] and fungal FAS [18,248] and results in an increased linker stiffness [8,252,253], which may prevent linker entanglement or guide motion paths of carrier proteins. The efficiency of substrate transfer by a carrier protein results from the interplay of its own motion with the conformational dynamics of its enzymatic target sites. Considerable conformational dynamics have been observed in most other carrier protein-based multienzyme systems, and are now well-characterized for Dra YCC (Figure 3.6A), PDC (Figure 3.6B) and animal FAS (Figure 3.6C) [250,254]. A

comparison of these systems reveals that in all of them domain motions on the 100 Å scale complement the mobility of carrier proteins, whose linkers have maximal lengths below 45 Å. Electron cryo-tomography demonstrates that in the *E. coli* PDC the E2 core stays rigid, while the surrounding E1 and E3 subunits, bound to the E2 core, adopt a variety of conformations, with a maximal range of motion of ~150 Å [20]. For animal FAS, cryo-electron microscopy confirmed a rotation, twisting and bending around a central hinge, which leads to domain displacement of more than 120 Å and an almost unrestrained motion of the C-terminal thioesterase domain, which is linked via a 30-residue linker to the acyl carrier protein [250].

Altogether, although many aspects of the interplay between catalysis and dynamics remain unknown, these data indicate a crucial and general role of large-scale dynamics in multienzyme systems employing tethered carrier proteins. Such systems include highly relevant assembly lines for the production of bioactive compounds and drug candidates, such as non-ribosomal polypeptide synthetases and PKS. Understanding and considering general principles of substrate transfer aided by structural dynamics in systems ranging from YCC to PKS is a crucial prerequisite for efficiently engineering versatile carrier protein-based biosynthetic factories.

3.5 Experimental Procedures

3.5.1 Protein expression and purification

Full-length (Genebank accession: NP_285633) and CT domain (residues 554-1091) of YCC were amplified from genomic Dra R1 DNA (DSMZ205309) by PCR and cloned into vector pAB1GN-his10, a Gateway® (Thermo Fisher Scientific) compatible derivative of the MultiBac™ insect cell expression plasmid pACEBACI (Geneva Biotech) coding for an additional N-terminal 10 x His-tag. Protein was overexpressed in Sf21 insect cells at 27°C according to the baculovirus expression system manual. Cells were lysed by sonication in 50 mM TrisHCl (pH: 7.5); 200 mM NaCl; 5 mM $MgCl_2$; 20 mM imidazole; 10% (w/v) glycerol; 5 mM β-mercaptoethanol (BME). Soluble protein was purified by metal affinity chromatography using a Ni-NTA column and elution with a linear gradient to 500 mM imidazole. Full-length YCC was diluted with dilution buffer (50 mM TrisHCl (pH: 7.5); 10% (w/v) glycerol; 5 mM BME) to 70 mM NaCl and subjected to anion exchange chromatography using PL-SAX 4000 Å resin (Agilent) and eluted with a linear gradient to 500 mM NaCl in the same buffer. Both, the full length Dra YCC and Dra YCC CT domain were subjected to size exclusion chromatography on a Superose 6 and Superdex 200 (GE Healthcare) column, respectively, equilibrated in crystallization buffer (30 mM TrisHCl (pH: 7.5); 150 mM NaCl; 5% (w/v) glycerol; 10 mM dithiothreitol (DTT)). Proteins were concentrated to 10 mg/ml using centrifuge concentrators (Millipore) and flash frozen in liquid nitrogen.

3.5.2 Protein characterization by SEC-MALS

For size exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) measurements, 20 µl or 100 µl samples of 4 mg/ml protein were applied to a GE Healthcare Superdex 200 5/150 GL SEC column equilibrated

overnight in crystallization buffer at 4 °C, using an Agilent 1100 series HPLC system. Light-scattering and differential refractive index measurements were made using Wyatt miniDawn TriStar detector and a Wyatt Optilab rRex detector, respectively. The inter-detector delay volumes, band broadening, and the light-scattering detector normalization, were calibrated according to the manufacturer's protocol using a 2 mg/ml BSA solution (Thermo Pierce) run in the same buffer. The absolute refractive index of the buffer was measured using the refractive index detector. The data were collected and processed using the Wyatt Astra software. The molar mass was calculated from a global fit of the light scattering signals from three detectors at different angles, and the differential refractive index signal, using algorithms in the Astra 5 software.

3.5.3 Limited proteolytic digest to obtain a stable BC domain

Full-length Dra YCC (10 mg/ml) was treated with 1:1000 subtilisin protease (Sigma), incubated for 2 h at 20 °C, and quenched with phenylmethylsulfonylfluoride. The BC containing fragment was purified on a Superdex 200 (GE Healthcare) column equilibrated in crystallization buffer. Protein was concentrated to 10 mg/ml using centrifugal concentrators (Millipore) and flash frozen in liquid nitrogen.

3.5.4 Crystallization and crystallographic data collection

The Dra YCC CT domain was crystallized at 25 °C using the sitting drop vapor diffusion method at 9 mg/ml and a reservoir (26% polyethylene glycol 3350; 0.08 M DL-malic acid) to protein ratio of 2:1 in a total drop volume of 1.5 µl. Crystals appeared after 7 days and continued to grow for 42 days to size of 130 x 130 x 25 µm. Crystals were transferred into cryoprotecant (reservoir solution containing

32% polyethylene glycol 3350 and 20% ethylene glycol) and vitrified in liquid nitrogen. BC domain was crystallized at 17 °C using the sitting drop vapor diffusion method at 10 mg/ml and a reservoir (0.03 M $MgCl_2$; 0.1 M HEPES (pH 7.5); 24% w/v poly(acrylic acid sodium salt) 5,100) to protein ratio of 1:1 in a total drop volume of 1.5 µl. Crystals appeared after 1 day and continued to grow for 7 days to a size of 170 x 170 x 100 µm. Crystals were transferred into cryoprotecant (reservoir solution containing 30% w/v Poly(acrylic acid sodium salt) 5,100 and 20% ethylene glycol) and vitrified in liquid nitrogen. All crystallographic data were collected at the Swiss Light Source (Paul Scherrer Institute) at beam line X06DA at 100 K using a Pilatus 2M detector (Dectris). Dra YCC CT data were collected at a wavelength of 1.0 Å with an exposure time of 0.25 seconds, a rotation angle of 0.25° and a detector distance of 0.245 m. YCC BC data were collected at a wavelength of 1.0 Å with an exposure time of 0.15 seconds, a rotation angle of 0.1° and a detector distance of 0.17999 m.

3.5.5 Structure determination of Dra YCC CT and twinned Dra YCC BC

Dra YCC CT data were processed using XDS [255,256]. Phases were obtained by molecular replacement using PHASER (CCP4 suite) [257-259]. The initial model was refined by iterative cycles of manual model building and real space refinement in *Coot* and refinement in phenix.refine [260-262]. The final model includes residues 574-1091 for chain A, 573-1091 for chain B, 572-1091 for chain C, 556-1091 as well as 7 residues of the TEV-cleavage site for chain D, 575-1091 for chain E and 572-1091 for chain F. Residues 1011-1032 show increased disorder and could only be modeled in chain B and D. Density around the loop from residue 720-730 is indicative of disorder or partial binding of a non-natural ligand. The Dra YCC BC data were processed using XDS. Molecular replacement was done using

Molrep [263]. The model was rebuilt using *Coot* and refined in REFMAC5 in iterative cycles [237]. Twin refinement, as implemented in REFMAC5, was used. MolProbity was used to assess final model quality [264]. The final model includes residues 2-464 for chain A and B. The B subdomain shows increased disorder and residues ranges 156-167 and 184-197 in chain A and residue ranges 159-168 and 183-194 in chain B have not been modeled. The loop 346-355 in chain A and 346-358 in chain B shows increased disorder or flexibility and could not be modeled.

3.5.6 Negative stain electron microscopy and image processing

Dra YCC samples used for EM were thawed on ice and diluted in crystallization buffer to 0.05 mg/ml. 200 µm copper grids were glow discharged for 20 s, sample was adsorbed for one minute and blotted using Whatman filters. The grid was washed three times using crystallization buffer, two times using H_2O and once using 2% uranyl acetate, followed by a 20 s staining with 2% uranyl acetate. Grids of Dra YCC were imaged on a CM-100 microscope (Philips) equipped with a Veleta 2k x 2k camera (Olympus) at 80 kV and a magnification of 130000 x. The pixel size was 0.37 x 0.37 nm. 9819 Particles were picked from 82 images using standard procedures in XMIPP [265]. All particles were normalized and bad particles from the semi-automated particle picking procedure were discarded manually. The remaining particles were windowed in 120 x 120 pixel images. After extraction, particles with a z-score of >3 were discarded and 9477 particles were aligned and classified into 24 2D class averages using maximum-likelihood target function in Fourier space (MLF2D) in 100 iterations [266]. Distances were measured manually from single picked particles out of top view 2D class averages in ImageJ [267]. Only distances of clearly visible BC domains were measured and

binned into classes of 5 Å increments. 2D projections of Dra CT domain were generated using EMAN2 [268].

3.5.7 Small angle X-ray scattering

Dra YCC was thawed on ice and dialyzed against 20 mM HEPES (pH 7.4), 150 mM NaCl, 5% Glycerol, 2 mM TCEP. Samples were measured at the Swiss Light Source (Paul Scherrer Institut) at the X12SA beamline at 3 mg/ml. Measurements were done in a 1 mm capillary at 10 positions, for 0.04 sec exposure per position and 10 acquisitions with 8 scan repeats. Data were processed, scaled and merged using the ATSAS package [244]. The ensemble analysis was done using ensemble optimization method (EOM) [243]. The three models contribute ~38, ~38 and ~25% to the ensemble and the model with highest contribution was selected for reconstitution of native-like linkers in *Coot*. Scattering curves of Dra YCC models were calculated using CRYSOL [269]. The BCCP model was generated using SWISS-MODEL [270].

3.5.8 Structure analysis

Interfaces of proteins were analyzed by PISA [271]; alignments of sequences were done using MUSCLE [272]. Active site of Dra YCC CT was analyzed using SiteMap [273,274]. Conservation was plotted using AL2CO, with the entropy-based algorithm [275] and figures were generated using The PyMOL Molecular Graphics System, Version 1.7 Schrödinger, LLC.

3.6 Author contributions:

Conceptualization, E.S. and T.M.; Investigation, A.H., M.H. and E.S.; Validation, A.H., M.H., E.S., and T.M.; Formal Analysis, A.H., M.H., E.S., and T.M.; Writing –

Original Draft, A.H., M.H., E.S., and T.M.; Writing – Review & Editing, A.H., M.H., E.S., and T.M.; Supervision: E.S. and T.M.; Project Administration: T.M.; Funding Acquisition: A.H., M.H., and T.M.

3.7 Acknowledgements

We acknowledge the staff of beamlines X06DA, X06SA and X12SA at Paul Scherrer Institute, Villigen, Switzerland, for support with crystallographic and small angle X-ray scattering data collection; the Biophysics facility of the Biozentrum for protein characterization and the University of Basel Center for Microscopy for electron microscopy. We acknowledge Christoph Sager for his help with active site characterization and docking of ligands. Moritz Hunkeler acknowledges generous support by a Novartis Excellence Fellowship. Anna Hagmann is supported by a “Fellowship for Excellence” of the Biozentrum Basel International PhD Program. This work was supported by Swiss National Science Foundation (SNF) project grants 138262, 159696 and SNF R’Equip grant 145023. The authors declare no conflict of interest.

3.8 Accession Numbers:

The atomic coordinates and structure factors have been deposited in the Protein Data Bank. www.pdb.org (Dra YCC BC: PDB ID code: 5H80; Dra YCC CT: PDB ID code: 5FIF).

3.9 Supplemental Data

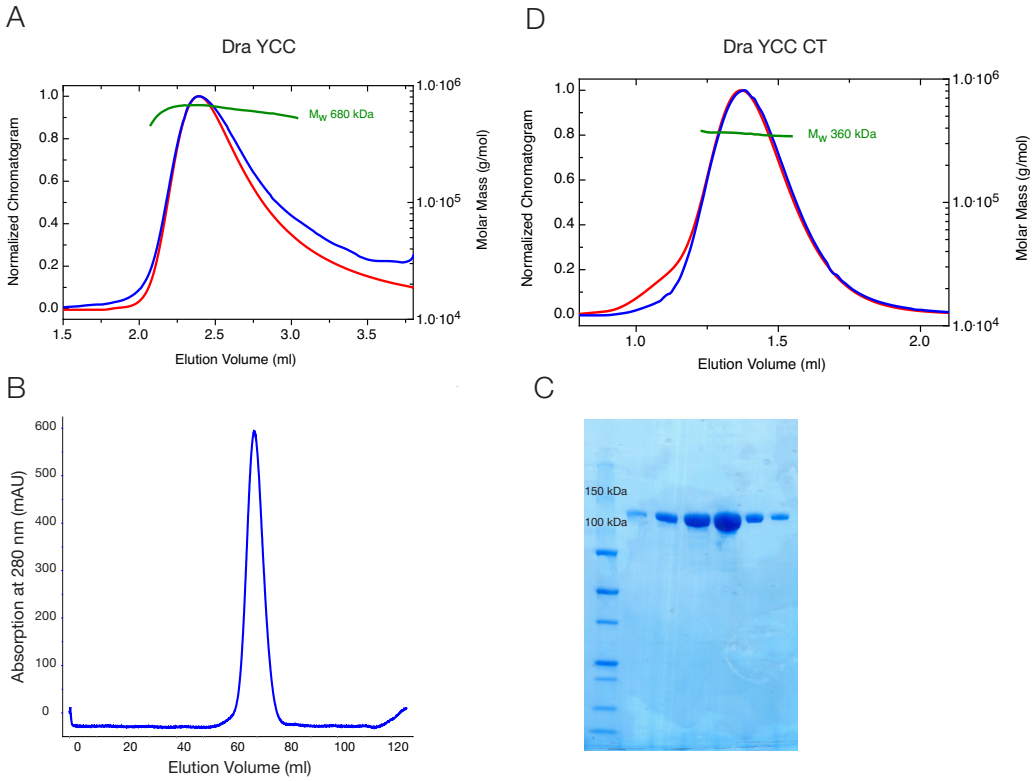
Supplemental Table 3.1 Crystallographic data collection and refinement statistics for Dra YCC CT and BC

	Dra YCC CT	**Dra YCC BC**
Wavelength (Å)	1.00003	1.0000
Resolution range (Å)	69.5-2.49 (2.58-2.49)	75.74-1.70 (1.8-1.70)
Space group	P 2_1 2_1 2_1	P 6_5
Unit cell (Å)	111.47, 149.57, 189.3	138.02, 138.02, 97.81
α, β, γ (°)	90, 90, 90	90, 90, 120
Total reflections	1,497,948 (237,650)	1,181,171 (180642)
Unique reflections	110,205 (17,213)	116,236 (18405)
Multiplicity	13.6 (13.8)	10.1 (9.8)
Completeness (%)	99.64 (97.4)	99.87 (97.8)
Mean I/sigma	13.34 (1.90)	15.61 (1.18)
Wilson B-factor ($Å^2$)	40.78	32.17
Twin operator/Twin fraction (%)	-	K, H, -L / 30.6
R-meas	0.22	0.10
CC 1/2	0.997 (0.676)	0.999 (0.461)
R-work	0.202	0.149
R-free	0.241	0.184
Number of atoms	46,326	7,346
Macromolecules	45,564	6,529
Ligands	80	28
Waters	682	789
Protein residues	3,089	856
RMS (°, angles)	0.83	2.46
RMS (Å, bonds)	0.0034	0.0262
Ramachandran favored (%)	96.3	96.3
Ramachandran outliers (%)	0.91	0.71
Clashscore	1.9	3.3

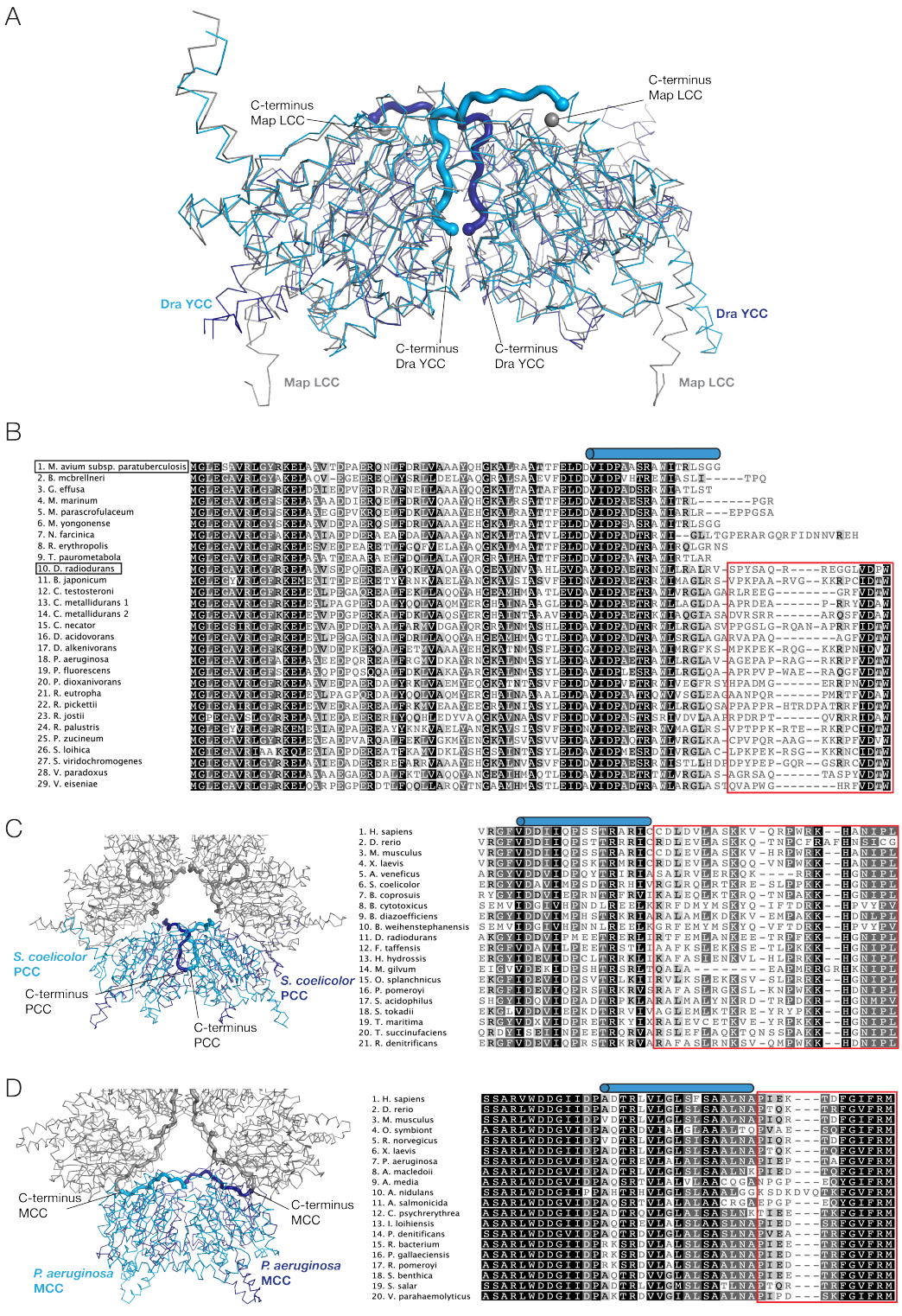
Values in parenthesis correspond to the highest resolution shell.

Supplemental Table 3.2 SAXS data statistics for Dra YCC

	Dra YCC
Instrument	cSAXS at SLS
Beam size (µm)	200 x 200
Wavelength (Å)	1.0000
Detector distance (m)	2.1368
Temperature (K)	293
Protein concentration (mg/ml)	3
Total exposure time (s)	240
Single exposure time (s)	0.04
Capillary diameter (mm)	1
I(0) (cm^{-1}) [from P(r)]	0.06-0.21
R_g (Å) [from P(r)]	84.31 ± 3.02
R_g (Å) (from Guinier)	80.03 ± 3.80
Porod volume estimate ($Å^3$)	1274980.00
I(0) (cm^{-1}) (from Guinier)	3.10 ± 1.09
D_{max} (Å)	256

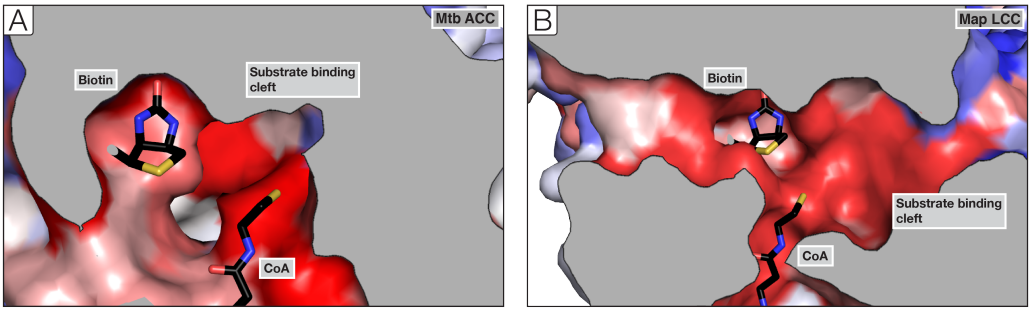


Supplemental Figure 3.1 SEC-MALS analysis of Dra YCC constructs and quality of full-length Dra YCC.
A. SEC-MALS analysis of full-length Dra YCC. Measured molecular weight of Dra YCC is 680 kDa. Raleigh ratio is shown in red, differential refractive index in blue. **B.** Size exclusion chromatogram of Dra YCC. Absorption at 280 nm is shown in blue, showing a symmetric peak, indicating a homogeneous sample. **C.** SDS-PAGE analysis of peak fractions of Dra YCC size exclusion chromatography showing purity of over 95%. **D.** SEC-MALS analysis of Dra YCC CT. Experimentally determined molecular weight of Dra YCC CT is 360 kDa. Raleigh ratio is shown in red, differential refractive index in blue. Figure related to Figure 3.1.

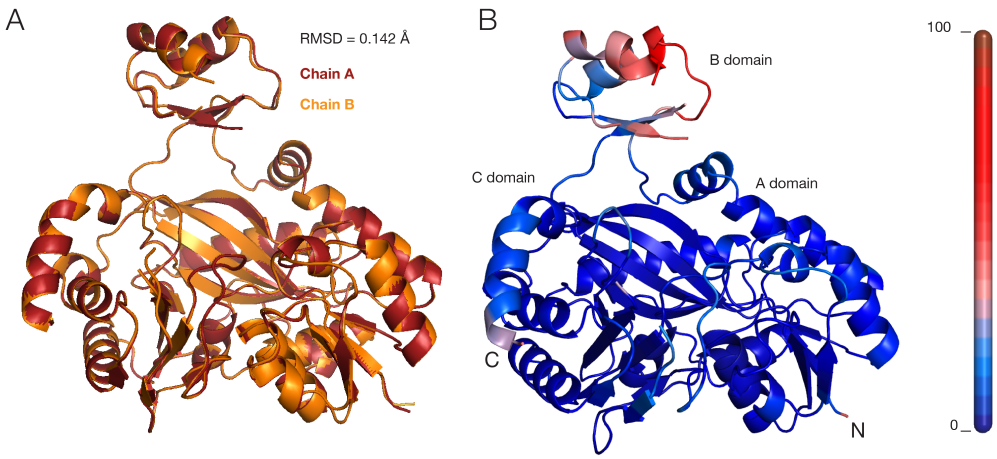


Supplemental Figure 3.2 C-terminal extension and alignment of 29 YCCs.
A. Superposition of Dra YCC (light and dark blue) onto Map LCC (gray). C-termini of both proteins are indicated. Relative to Map LCC, Dra YCC contains a C-terminal extension shown as bold loop. **B.** Alignment of 29 YCC sequences shaded according to identity level. C-terminal core fold helices in Dra YCC CT and Map LCC structures are indicated in blue, the C-terminal extension with conserved DXW motif is indicated in red. **C.** Left panel: Ribbon representation of *S. coelicolor* PCC (Arbolaza et al., 2010). Right panel: Alignment of 21 PCC sequences. **D.** Left panel: Ribbon representation of *P. aeruginosa* MCC (Huang et al., 2011)). Right panel: Alignment of 20 MCC sequences. In C) and D), C-

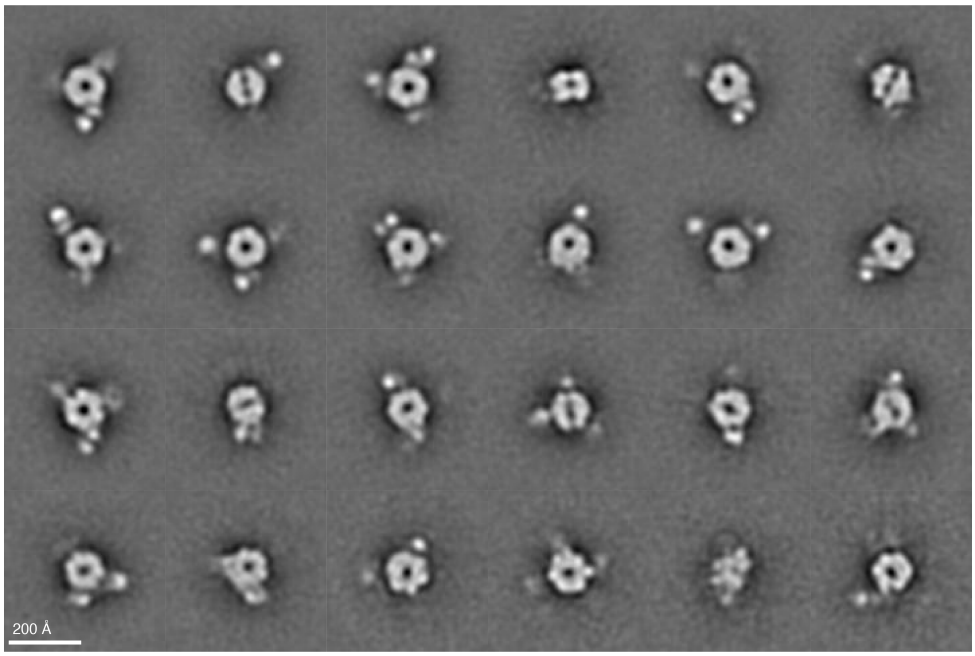
terminal extensions are shown as bold loops. The C-terminal core fold helix is indicated in blue; further C-terminal extensions are indicated in red. Sequence alignments are shaded according to identity level. Figure related to Figure 3.1.



Supplemental Figure 3.3 Active site cleft of ACC and LCC.
Surface representation of the active site of **A.** Mtb ACC and **B.** Mal LCC[276,277]. Coloring according to conservation from blue to red. Biotin and CoA moiety are modeled according to their positioning in the MCC holoenzyme (PDB ID: 3U9R[228]). Figure related to Figure 3.2.



Supplemental Figure 3.4 BC protomers in the asymmetric unit and B-factors
A. Overlay of the two BC protomers of the asymmetric unit. One protomer is shown in red, the other in orange. RMSD of 0.142 Å is indicated. **B.** B-factor plot of one BC protomer. Scale bar from 0 to 100 is shown going from dark blue to dark red. Increased B-factor values in the B-domain can be observed, likely due to the absence of a bound substrate. Figure related to Figure 3.3.



Supplemental Figure 3.5 Negative stain EM 2D class averages of Dra YCC.
9477 particles were classified into 24 2D class averages using the MLF2D method (Scheres et al., 2007). Classes are shown with decreasing number of particles per class, from top left (592 particles) to bottom right (179 particles). Scale bar is 200Å. Figure related to Figure 3.4.

4 Dimeric Structure of the Bacterial Extracellular Foldase PrsA

This research was originally published in Journal of Biological Chemistry.

Roman P. Jakob, Johanna R. Koch, Björn M. Burmann, Philipp A. M. Schmidpeter, Moritz Hunkeler, Sebastian Hiller, Franz X. Schmid, Timm Maier.
Dimeric Structure of the Bacterial Extracellular Foldase PrsA.
J Biol Chem. 2015; **290(6):** 3278-3292.

4.1 Abstract

Secretion of proteins into the membrane-cell wall space is essential for cell wall biosynthesis and pathogenicity in Gram-positive bacteria. Folding and maturation of many secreted proteins depend on a single extracellular foldase, the PrsA protein. PrsA is a 30 kDa protein, lipid-anchored to the outer leaflet of the cell membrane. The crystal structure of *Bacillus subtilis* PrsA reveals a central catalytic parvulin-type prolyl isomerase domain, which is inserted into a larger composite NC domain formed by the N- and C-terminal regions. This domain architecture resembles, despite a lack of sequence conservation, both trigger factor, a ribosome-binding bacterial chaperone, and SurA, a periplasmic chaperone in Gram-negative bacteria. Two main structural differences are observed in that the N-terminal arm of PrsA is substantially shortened relative to trigger factor and SurA and in that PrsA is found to dimerize in a unique fashion via its NC domain. Dimerization leads to a large, bowl-shaped crevice, which might be involved *in vivo* in protecting substrate proteins from aggregation. NMR experiments reveal a direct, dynamic interaction of both the parvulin and the NC domain with secretion propeptides, which have been implicated in substrate targeting to PrsA.

4.2 Introduction

Protein folding is assisted by molecular chaperones and by foldases. Peptidyl-prolyl *cis-trans* isomerases (PPIases) constitute a class of foldases that occur in all types of cells and cell compartments[155]. They catalyze the isomerization of peptide bonds preceding proline residues[158], which is often a rate-limiting step during protein folding[278]. Three families of PPIases are known: the cyclophilins[279], the FK506-binding proteins (FKBP)[280], and the parvulins[157]. Parvulins are ubiquitous globular proteins or protein domains of about 100 residues characterized by the parvulin fold[161], a four-stranded antiparallel β-sheet, surrounded by four α-helices ($\beta\alpha_3\beta\alpha\beta_2$). Human Pin1 is the most prominent member of the parvulin family and is involved in phosphorylation-dependent signal transduction pathways, in transcriptional regulation, and cell cycle control[281]. In prokaryotes, parvulins assist in the maturation of intracellular proteins, e.g. nitrogenase reductase[282] and various virulence factors[283,284]. A network of periplasmic PPIases, including SurA, FkpA, CypB and Par10, supports the folding of periplasmic and outer membrane proteins in Gram-negative bacteria[166,167]. In Gram-positive bacteria, the parvulin-type PPIase PrsA is the only general factor mediating folding of secreted proteins, which are essential for bacterial pathogenicity and cell wall biosynthesis.

PrsA is a ubiquitous 30 kDa lipoprotein localized to the space between plasma membrane and cell wall. It is tethered to the outer leaflet of the cell membrane by a lipid anchor, which is attached to its N-terminal cysteine residue[168,169]. PrsA is essential under normal growth conditions in *B. subtilis* [170] and PrsA depleted cells are affected by decreased cell wall integrity[172,173], osmotic shock susceptibility[171] and increased sensitivity to antibiotics[174,175]. Overexpression of PrsA improves the recombinant overproduction of biotechnologically important proteins[163,285]. PrsA plays an important role as a folding factor of secreted proteins[286] including enzymes involved in cell wall biogenesis[173], toxins[176], and virulence factors[174,177]. Due to its general role for the maturation of pathogenicity factors, PrsA is a potential target for novel antimicrobial drugs.

At the sequence level, PrsA consists of three regions: a large N-terminal part with unknown function, followed by a parvulin-type PPIase domain[287] and a small C-terminal region[169]. Despite its importance for protein secretion and pathogenicity in Gram-positives, the structural basis for PrsA function has so far remained

unknown. Here, we determine the crystal structure of PrsA from *B. subtilis* and characterize its functional properties. Our data reveal a didomain architecture, in which the catalytically active parvulin domain is inserted into a composite N- and C-terminal chaperone-like domain. Dimerization of a soluble PrsA variant at high concentrations and presumably of membrane-tethered PrsA is mediated by the chaperone-like domain and creates a bowl-shaped crevice, which may be required for *in vivo* chaperone activity of PrsA. NMR spectroscopic experiments provide initial insights to substrate interactions of PrsA.

4.3 Experimental Procedures

4.3.1 Expression and purification of PrsA

For the expression of full length PrsA (residues 21–292) and C-terminal truncated version (residues 21–280) the gene fragments were amplified by PCR from *B. subtilis* (strain 168) and cloned into the expression plasmid pNIC28-Bsa4[288], where they are N-terminally linked to hexa-Histag followed by a TEV-site. The proteins were overexpressed in *E. coli* BL21(DE3). After lysis of the cells in 50 mM Hepes/NaOH, 500 mM NaCl, pH 7.4, 40 mM imidazol with a sonicator and centrifugation, both proteins were found in soluble form. The proteins were purified by immobilized metal-affinity chromatography on a Ni-NTA column, TEV digested overnight followed by an orthogonal Ni-NTA column step and then subjected to size exclusion chromatography in 20 mM Hepes/NaOH pH 7.4 on a Superdex S75 column (GE Healthcare). The protein-containing fractions were pooled and concentrated in Amicon Ultra units (Millipore). [*U*-^{15}N]-PrsA was obtained by growing *E. coli* BL21(DE3) cells in M9 minimal medium supplemented with 1g/l ($^{15}NH_4$)Cl (Cambridge Isotope Labs) at 37 °C until an OD_{600} = 0.8 was reached. Expression was induced by addition of 1 mM isopropyl-β-D-thiogalacto-pyranosid (IPTG). Se-Methionine-derivatized PrsA was produced in M9 minimal medium. Cells were grown at 37 °C. At an OD_{600} of 0.6 100 mg/L lysine, 100 mg/L Phenylalanine, 100 mg/L threonine, 50 mg/L isoleucine, 50 mg/L leucine, 50 mg/L valine, 100 mg/L DL-Selenomethionine were added, and after 30 min of incubation, protein expression was induced by addition of 1 mM IPTG. Cells were harvested 6 h after induction and purified in the same manner as the unlabelled protein. Yields were about 15 – 30 mg/l. RCM-T1 was expressed and purified as described[289].

4.3.2 Prolyl isomerase and chaperone activity assays

The prolyl isomerase activities were measured by a protease-free fluorescence assay, as described[290]. The kinetics of the decrease in *cis*-content was measured by the change in fluorescence at 416 nm (5 nm bandwidth) after excitation at 316 nm (3 nm bandwidth) in 100 mM K-phosphate 7.4 at 15 °C. Under these conditions, the *cis*-to-*trans* isomerization of the prolyl bond was a mono-exponential process, and its rate constant was determined by using GraFit 5.0 (Erithacus Software, Staines, UK). The folding experiments of RCM-T1 were performed as described[289]. The citrate synthase aggregation assay was measured as described[291]. Citrate synthase was unfolded in 50 mM Tris–HCl (pH 8.0), 20 mM dithioerythritol, 6 M GdmCl for 1 h and then diluted 200-fold to a final concentration of 0.15 µM (monomer) in 50 mM Tris–HCl (pH 8.0), 0.1 mM dithioerythritol, 30 mM GdmCl, and various concentrations of prolyl isomerase at 25 °C. Spontaneous aggregation was monitored by measuring the increase in light-scattering at 360 nm.

4.3.3 Thermal induced unfolding transitions

The heat-induced unfolding transitions were measured by CD at protein concentrations of 0.5 – 4.0 µM in 100 mM K phosphate (pH 7.4). Samples were heated at a rate of 60 K h^{-1}. The transitions were monitored by the increase of the CD signal at 222 nm with 1 nm bandwidth and 10 mm path length with a JASCO J710A spectropolarimeter (JASCO, Tokyo, Japan). The experimental data were analyzed on the basis of a three-state approximation with a fixed heat capacity change ΔC_p of 6000 J mol^{-1} K^{-1} for the first transition and a ΔC_p of 6000 J mol^{-1} K^{-1} for the second transition[292,293].

4.3.4 NMR spectroscopy

NMR experiments of PrsA were performed in NMR buffer containing 25 mM MES, 150 mM NaCl, pH 6.5. The measurements were recorded at 298 K on a Bruker AscendII 700 MHz spectrometer equipped with a cryogenically cooled triple-resonance probe. For the titration with AmyE–(ETANKSNELTA)–peptide 2D [^{15}N,^{1}H]-TROSY-HSQC[294] spectra of 0.5 mM [*U*-95% ^{2}H,^{15}N]–PrsA at different peptide concentration titrated from a 50 mM stock solution were collected. NMR

data were processed using PROSA[295] and analyzed with CARA[296]. The chemical shift changes of the amide moiety were normalized according to equation 1.

$$\Delta\delta\ i\ \sqrt{(\Delta\delta 2^{1}H))^2 + 0.2\ (\Delta\delta 2^{1\cdot}N))^2} \quad (1)$$

The chemical shift changes observed upon increasing AmyE-peptide concentrations were analyzed by non-linear regression to equation 2,

$$\Delta\delta_{o2s}\ i\ \Delta\delta_{max}\frac{(K_D+[peptide]_0+[PrsA]_0)-\sqrt{2K_D+[peptide]_0+[PrsA]_0)^2-24[PrsA]_0[peptide]_0)}}{2[PrsA]_0} \quad (2)$$

by using standard software.

4.3.5 SEC-MALS

For size exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) measurements, 20 µl or 100 µl samples of 1-4 mg/ml protein were applied to a GE Healthcare Superdex 200 5/150 GL SEC column equilibrated overnight in 20 mM Hepes pH 7.4 at 5 °C, using an Agilent 1100 series HPLC system. Light-scattering and differential refractive index measurements were made using Wyatt miniDawn TriStar detector and a Wyatt Optilab rRex detector, respectively. The inter-detector delay volumes, band broadening, and the light-scattering detector normalization, were calibrated according to the manufacturer's protocol using a 2 mg/ml BSA solution (Thermo Pierce) run in the same buffer. The absolute refractive index of the buffer was measured using the refractive index detector. The data were collected and processed using the Wyatt Astra 5 software. The molar mass was calculated from a global fit of the light scattering signals from three detectors at different angles, and the differential refractive index signal, using algorithms in the Astra 5 software.

4.3.6 Protein crystallization and structure determination

Initial full-length PrsA crystals diffracted only to 6 Å resolution. After truncation of the twelve C-terminal residues (SNSTSSSSSNSK) PrsAΔC crystallized under identical conditions as the full-length protein, but crystals diffracted to higher resolution. Best diffracting crystals were grown at 30 °C in 17% PEG2000MME, 0.1 M Hepes-NaOH pH 7.5 within 4 days. The crystals were dehydrated and cryo-preserved by successively increasing the PEG2000MME to 50%. Diffraction

data were collected at Swiss Light source X06DA beamline and processed using XDS[200]. The crystals belong to space group $C222_1$ with cell dimensions *a* = 71.5 Å, *b* = 88.4 Å, and *c* = 233.4 Å, α=β=γ= 90° and contain two molecules per asymmetric unit. The structure of PrsA was determined by SeMet-SAD phasing and refined to 2.63 Å. All expected 2 x 4 selenium sites were identified in the Se-Met dataset by Phenix-HySS[297]. Initial automated model building was carried out with Phenix[298] and Buccaneer[299]. Manual model building and structure refinement were performed with Coot[300] and PHENIX[301]. The final model includes residues 3–259 of mature PrsA.

4.4 Results

4.4.1 PrsA is a stable two-domain protein that dimerizes at high concentrations

In Gram-positive bacteria, PrsA-like lipoproteins have an N-terminal cysteine that is enzymatically modified upon signal sequence cleavage with a diacylglycerol residue to promote membrane association[302]. We produced a recombinant soluble variant of the PrsA protein that comprises residues 21–292 excluding the signal peptide and the aminoterminal residue Cys20. The circular dichroism spectrum shows a maximum near 190 nm and minima near 208 and 222 nm, as expected for a protein with a high content of helical secondary structure (Figure 4.1A). Thermal unfolding of PrsA was monitored via the increase of the helical CD signal at 222 nm. The resulting curve can be decomposed into two transitions with midpoints (T_{M1}) at 46.5°C and (T_{M2}) at 62.8°C and unfolding enthalpies (ΔH_D) of 247 kJ mol^{-1} and 211 kJ mol^{-1}, respectively (Figure 4.1B). To distinguish whether these transitions involve an oligomerization process, we set out to characterize the oligomeric state of soluble PrsA as a function of the protein concentration. The two thermal unfolding transitions are independent of the protein concentration in the range between 0.5 and 4 µM (Figure 4.1C). Furthermore, at a concentration of about 25 µM, PrsA elutes in a single peak with monomer molecular weight from size exclusion chromatography coupled to multiangle light scattering (SEC-MALS) (Figure 4.2A). These data thus indicate that PrsA is monomeric in the concentration range up to at least 25 µM and that the two concentration-independent transitions observed during thermal unfolding

(Figure 4.1B) represent conformational unfolding reactions of individual domains of monomeric PrsA.

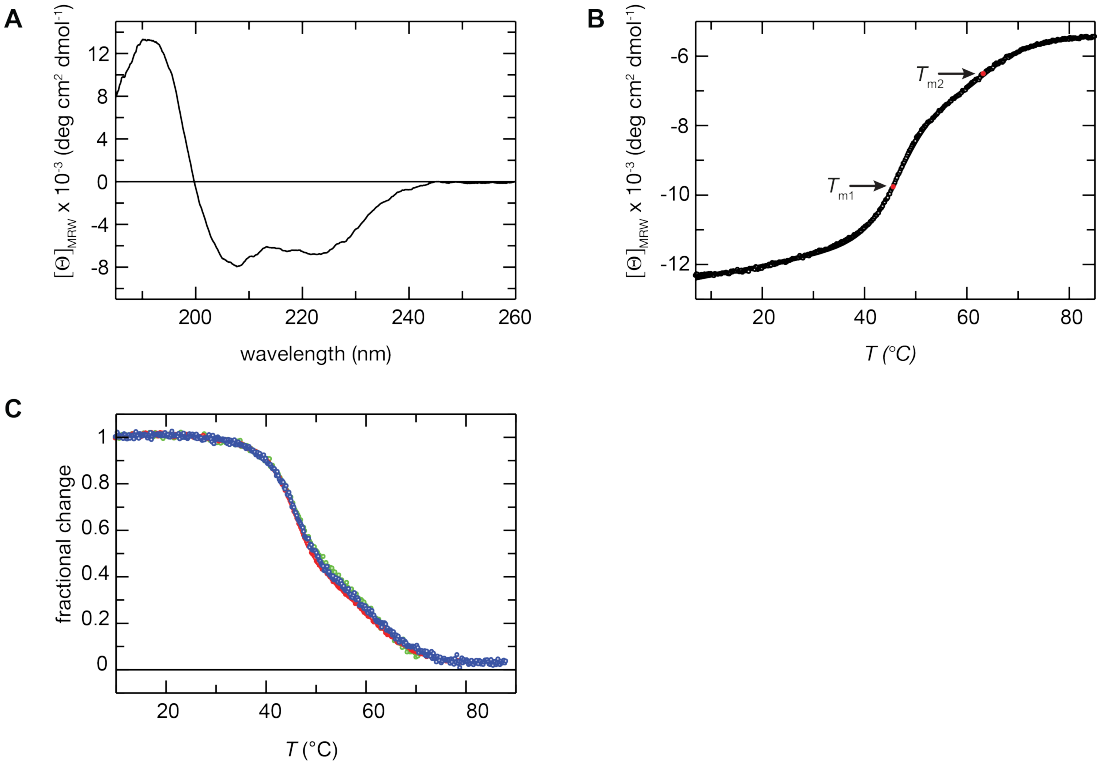


Figure 4.1 Circular dichroism spectra and thermal stability of PrsA.
A Far-UV CD spectrum of PrsA at 15°C. The spectrum was measured with 2 µM protein in 10 mM potassium phosphate (pH 7.0), at a path length of 1 mm and a band width of 1 nm. **B** Thermal stability of PrsA. Thermal transition measured by CD at 222 nm for the wild-type protein. The change of the CD signal is shown as a function of temperature. The three-state analysis (continuous line) resulted in T_{M1} = 46.5 ± 0.4°C, ΔH_{D1} = 247 ± 15 kJ mol^{-1} and T_{M2} = 62.8 ± 0.4°C, ΔHD2 = 211 ± 15 kJ mol^{-1}. T_{M1} and T_{M2} are indicated as red dots. The transition was measured with 2 µM protein in 100 mM potassium phosphate (pH 7.0), at a path length of 10 mm. **C** Thermal stability of PrsA at different protein concentrations. The fractional change after a three-state analysis is shown as a function of the temperature. The transitions were measured with 0.5 µM (red; T_{M1} = 46.7 ± 0.2°C, ΔH_{D1} = 259 ± 12 kJ mol^{-1}, T_{M2} = 62.4 ± 0.2°C, ΔH_{D2} = 201 ± 13 kJ mol^{-1}), 2 µM (green) and 4 µM (blue; T_{M1} = 46.2 ± 0.3 °C, ΔH_{D1} = 233 ± 14 kJ mol^{-1}, T_{M2} = 67.0 ± 0.3 °C, ΔH_{D2} = 218 ± 18 kJ mol^{-1}) protein in 100 mM potassium phosphate (pH 7.0), at a path length of 10 mm.

While SEC-MALS is an ideal tool to study oligomeric protein states at intermediate and low concentrations, it is not suited for measurements at high protein concentrations. However, increased concentrations best represent the *in vivo* situation of a membrane-anchored protein, where high local concentrations and molecular pre-orientation increases the likelihood of dimer formation by up to 10^6-fold[303]. Thus, PrsA oligomerization was also analyzed at high protein concentrations by lateral diffusion and relaxation analyses of ^{15}N-labelled protein

by NMR spectroscopy. The TRACT and BPP–LED NMR experiments demonstrate that at 750 µM PrsA is dimeric (Figure 4.2B, C), pointing to an inherent propensity of PrsA to dimerize based on direct interactions of its soluble regions. Lipid anchoring will further promote self-association, indicating that a dimer is the most likely physiological state of PrsA, in agreement with earlier *in vivo* data[173,174].

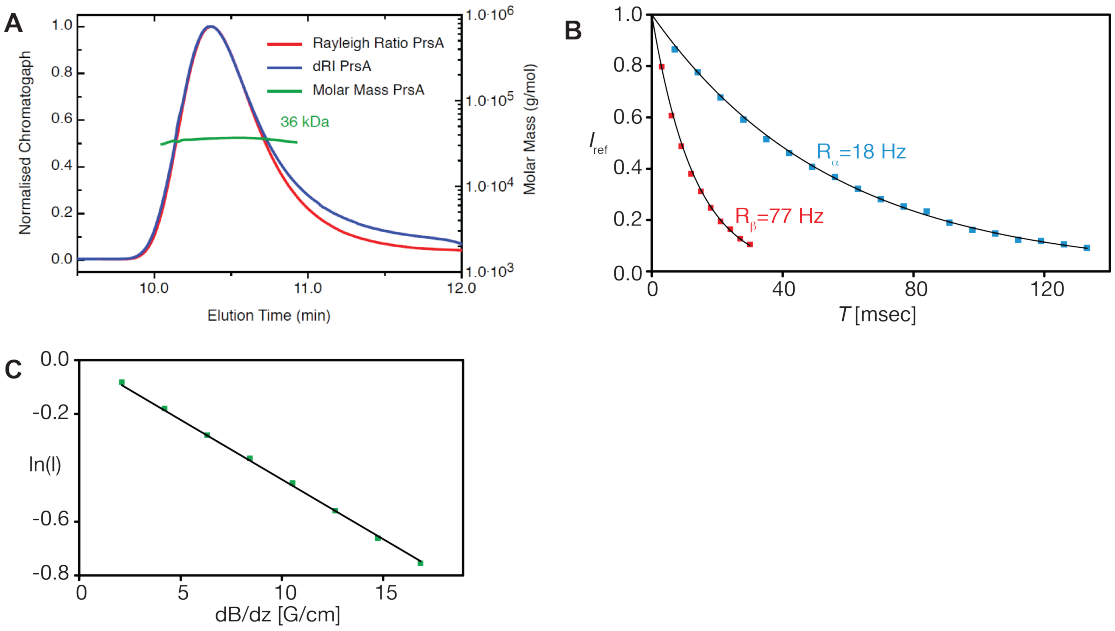


Figure 4.2 Molecular weight and biophysical molecular parameters of PrsA.
A. Size exclusion chromatography coupled to static multi angle light scattering and refractive index measurements (SEC-MALS) of PrsA. The molecular mass was calculated throughout the eluting peaks and is indicated in green. The calculated mass of 36.0 kDa is in good agreement with the 33.4 kDa for N-terminal His and TEV tagged (MHHHHHHSSGVDLGTENLYFQ*SM (*, TEV cleavage site) PrsA protein (residues 21–292). **B.** [^{15}N,^{1}H]-TRACT experiment[304] for the determination of the effective rotational correlation time constant τc. The 1D proton signal intensity, I_{ref}, of a sample of 750 µM PrsA at 25°C was integrated between 10 and 8.5 ppm and plotted vs. the relaxation period *T*. The transverse relaxation rates of the TROSY component (blue; black: least-squares fit) and the anti-TROSY component (red; black: least-squares fit) were determined to be R_α = 18 Hz and R_β = 77 Hz, resulting in τc = 23 ns, that is typical for a globular protein of about 60 kDa size[305]. **C.** Measurement of the molecular diffusion constant in aqueous solution by the ^{15}N-filtered diffusion BPP–LED NMR experiment[306] for 750 µM PrsA (green). The logarithm of the signal intensity, integrated between 10 and 8.5 ppm, is plotted vs. the gradient strength of the pulsed field gradients. The black line represents the linear fit to the measured data. This measurement yielded the self-diffusion constants D_0 = 6.11*10-11 m^2s^{-1}, that corresponds under the assumption of a spherical molecule and by using the Stokes-Einstein-equation to an estimated molecular weight of 58 kDa.

4.4.2 Catalysis of prolyl isomerization by PrsA in peptides and proteins

To characterize the activity of PrsA as a prolyl isomerase and to map its substrate specificity, we used a fluorimetric protease-free assay and a set of proline-containing tetrapeptides, which carry an aminobenzoyl (Abz) group at the aminoterminus and a para-nitroanilide (pNA) group at the carboxyterminus[290,307].

In the peptides of the general formula Abz-Ala-Xaa-Pro-Phe-pNA, the Xaa position was occupied by either charged (Glu, Lys), aliphatic (Ala, Leu), or aromatic (Phe) residues. PrsA catalyzes peptide isomerization very well when hydrophobic residues such as Leu precede the proline (Abz-Ala-Leu-Pro-Phe-pNA), and 16 nM PrsA sufficed to double the isomerization rate (Figure 4.3A). From the increase of the isomerization rate as a function of the PrsA concentration (Figure 4.3C), the catalytic efficiencies (k_{cat}/K_M) of PrsA towards peptides with six different Xaa-Pro sequences were derived (Table 4.1).

Table 4.1 Catalytic efficiencies of *Bacillus subtilis* PrsA for prolyl isomerization in peptide and protein substrates

	BsPrsA[a]	BsPrsA[b]	SaPrsA(140–245)[c]
Peptide isomerization	k_{cat}/K_M ($mM^{-1}s^{-1}$)	k_{cat}/K_M ($mM^{-1}s^{-1}$)	k_{cat}/K_M ($mM^{-1}s^{-1}$)
Abz-Ala-**Ala**-Pro-Phe-pNA	52	6	17
Abz-Ala-**Glu**-Pro-Phe-pNA	450	7	33
Abz-Ala-**Leu**-Pro-Phe-pNA	$3 \cdot 10^3$	n.d.[d]	n.d.
Abz-Ala-**Lys**-Pro-Phe-pNA	300	15	5
Abz-Ala-**Phe**-Pro-Phe-pNA	500	n.d.	n.d.
Abz-Ala-**Asn**-Pro-Phe-pNA	n.d.	0.5	3
Protein folding			
RNase T1	8.7[e]	25[f]	n.d.

[a] Catalytic activities towards peptides of full-length PrsA from *B. subtilis* from protease-free experiments as described in Figure 4.3. The confidence limits of the k_{cat}/K_M values were between 5% and 10%.

[b] Catalytic activities towards peptides of full-length PrsA from *B. subtilis* derived with a protease-coupled assay. Data taken from [169,308].

[c] Catalytic activities of the isolated parvulin domain of *S. aureus* SaPrsA(140–245). Data taken from [169,308].

[d] n.d., not determined.

[e] Catalytic activity was determined using the refolding of reduced and carboxymethylated RNase T1 (RCM-T1) in 2 M NaCl, 100 mM Tris-HCl, pH 7.8, and 1 mM EDTA.

[f] Catalytic activity was determined using the refolding of urea denaturated RNase T1 in 0.28 M urea, 100 mM Tris-HCl, pH 7.8, and 1 mM EDTA.

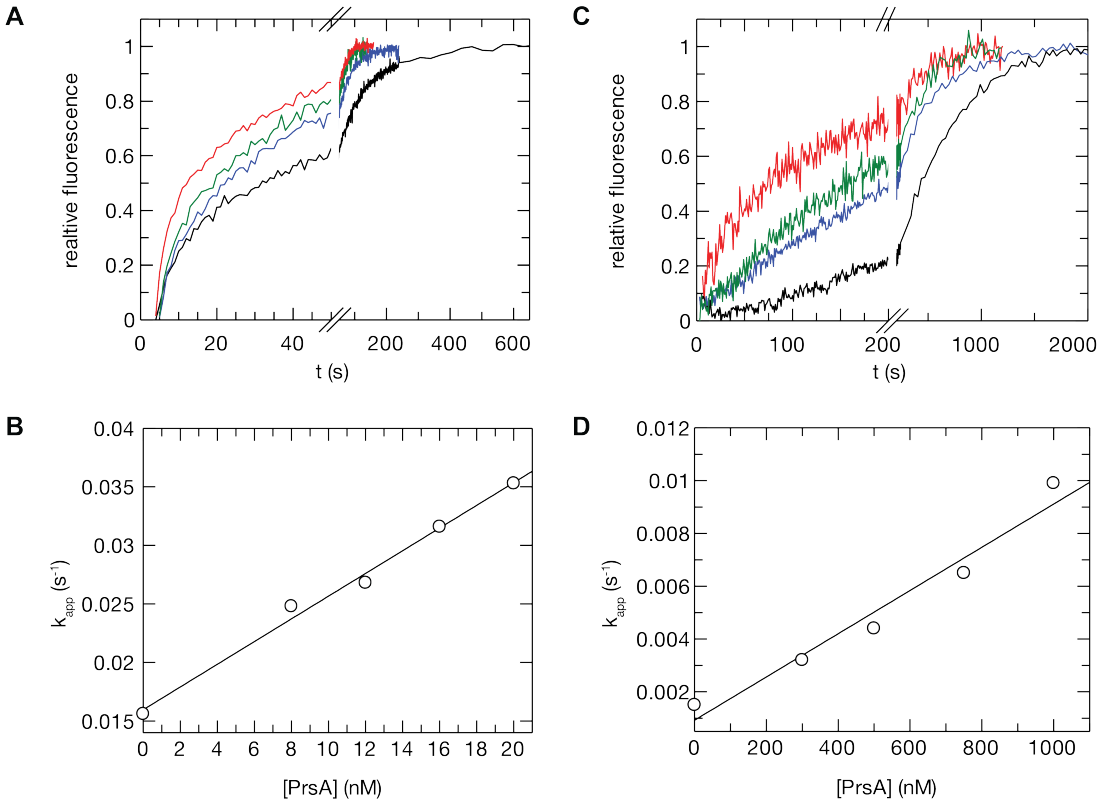


Figure 4.3 Characterization of enzymatic activity of PrsA.
A. Kinetics of *cis-trans* isomerization of 3 µM Abz-Ala-Leu-Pro-Phe-pNa followed by fluorescence at 416 nm, without enzyme (black), with 12 nM (blue), 16 nM (green) and 20 nM (red) PrsA. **B.** Refolding kinetics of RCM-T1 in the presence of increasing concentrations of PrsA, 0 nM (black), 300 nM (blue), 500 nM (green) and 1000 nM (red). Catalytic efficiencies of PrsA for **C.** the *cis-trans* isomerization of Abz-Ala-Leu-Pro-Phe-pNA and **D.** the refolding of RCM-T1. The measured rate constants k_{app} are shown as a function of the prolyl isomerase concentration. The k_{cat}/K_M values derived from the slopes are given in Table 4.1.

The highest catalytic efficiency was observed for Leu-Pro. For Phe-Pro the efficiency was 6-fold, for Ala-Pro, 60-fold lower. Peptides with charged residues such as Glu or Lys preceding the proline resembled the peptide with the Phe-Pro sequence in their k_{cat}/K_M values (Table 4.1). Our k_{cat}/K_M values are about 10-to 20-fold higher than those obtained previously for PrsA with a protease-coupled activity assay using N-terminally succinylated Suc-Ala-Xaa-Pro-pNA peptides (Table 4.1)[169]. The protease-coupled peptide assay can be used only for enzymes that are resistant to proteolysis, and therefore is unsuitable for most multi-domain prolyl isomerases. Possibly, PrsA is sensitive to proteolysis and inactivated during the assay as found for trigger factor[309] and SlyD[310]. With its k_{cat}/K_M value of 3×10^3 $mM^{-1}s^{-1}$ for the Leu-Pro peptide, PrsA shows an activity that is 1000-fold and 10-fold higher than the activities of human parvulin14 and *E. coli* SurA, respectively, and only 10-fold lower than the activity of *E. coli* parvulin 10, which

is the most active parvulin known to date[311]. In its substrate specificity with preference for hydrophobic residues preceding proline, PrsA resembles other bacterial parvulin homologues[312,313].

The efficiency of PrsA as a catalyst of proline-limited protein folding (the protein folding activity) was determined by using reduced and carboxymethylated ribonuclease T1 (RCM-T1) as a substrate protein. This variant cannot form disulfide bonds because the cysteine residues are modified. RCM-T1 is permanently unfolded in the absence of salt, but refolds spontaneously in 2 M NaCl, where electrostatic repulsion between negatively charged groups is decreased. In unfolded proteins, most of the proline residues adopt the *trans* conformation. Proline 39 is in *cis* conformation in folded RCM-T1, but in the unfolded state 85% of the molecules contain a trans-proline 39. As a consequence, the refolding of 85% of all molecules is limited in rate by *trans→cis* isomerization at Pro39, which shows a time constant of 530 s (at 15 °C, pH 8.0). The major part of the fluorescence change during RCM-T1 refolding reflects this isomerization and makes RCM-T1 an excellent substrate for assaying the folding activity PPIases[314-317]. Relatively high PrsA concentrations were necessary to accelerate the refolding of RCM-T1 (Figure 4.3B). The apparent rate constant of catalyzed folding increased in a linear fashion with PrsA concentration suggesting that the affinity of PrsA for refolding RCM-T1 is rather low (Figure 4.3D). A k_{cat}/K_M value of 8.7 $mM^{-1}s^{-1}$ is derived from Figure 4.3D. This value is similar to previously determined values (Table 4.1)[169] and also similar to the k_{cat}/K_M values obtained for the RCM-T1 refolding catalyzed by parvulin10 or SurA from *E. coli*[311,318]. In its properties as a catalyst of protein folding, PrsA thus resembles other parvulins.

4.4.3 Crystal structure of PrsA

PrsA (21–292) crystallized under various conditions, but diffraction was always limited to 6 Å resolution. At their C-termini, many PrsA orthologs contain serine-rich stretches of five to fifteen residues. In PrsA, this sequence region was shown to be dispensable for function, *in vivo* and *in vitro*[169]. The targeted deletion of the respective twelve C-terminal residues (SNSTSSSSSNSK) from PrsA improved diffraction to about 3.5 Å resolution under similar conditions as for the full-length protein. Crystal dehydration by successive addition of long chain polyethylene

glycols (up to 50% PEG 2000 MME) led to a further improvement. Finally, the X-ray crystal structure of PrsA (21–280) was solved by single-wavelength anomalous dispersion (SAD) for crystals of Se-Methionine-derivatized PrsA at 3.2 Å resolution. The structure of native PrsA was refined to R/R_{free} of 22/25% at 2.63 Å resolution. Data collection and refinement statistics are given in Supplemental Table 4.1.

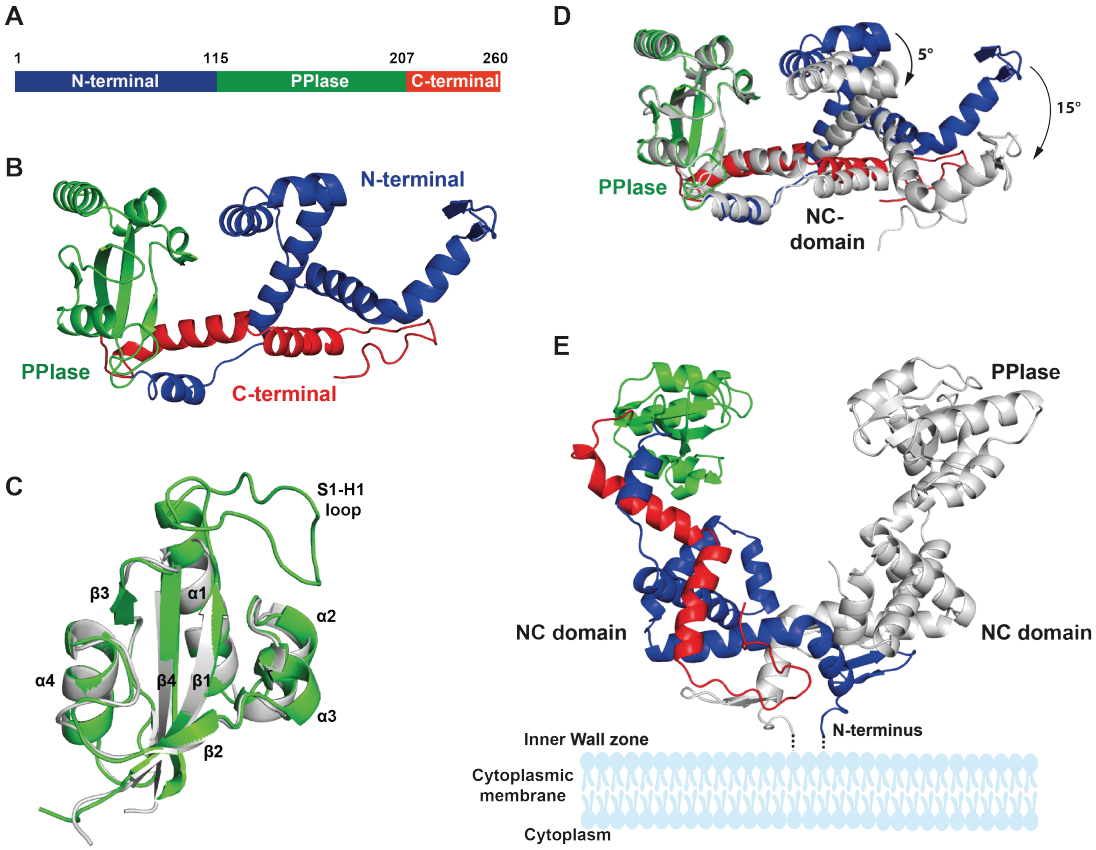


Figure 4.4 Crystal structure of PrsA.
A. Schematic representation of the PrsA primary structure. **B.** PrsA monomer. The N-terminal domain, PPIase and C-terminal domain are colored in blue, green and red, respectively. **C.** Superimposition of the parvulin domain of PrsA (gray) to the closest homologue human Pin1 (1pin, green). Secondary structure elements as well as the substrate interacting loop (S1-H1) of Pin1 are indicated. **D.** Comparison of the two PrsA molecules A and B in the asymmetric unit. PrsA molecule A is colored as in (B). PrsA molecule B is shown in gray, superimposed on molecule A based on the parvulin domain. The tilt of secondary elements is indicated. **E.** Dimeric structure of PrsA. PrsA molecules A and B are colored as in (D). The estimated position of membrane and membrane anchor are indicated in schematic representation.

PrsA consists of two domains in a discontinuous arrangement (Figure 4.4A): The central parvulin domain (Leu115–Glu207) (Figure 4.4B) is connected via short linkers to a composite domain formed by the N-terminal region Ser4–Gly114 together with the C-terminal region Arg208–Ser260. This domain, which we call

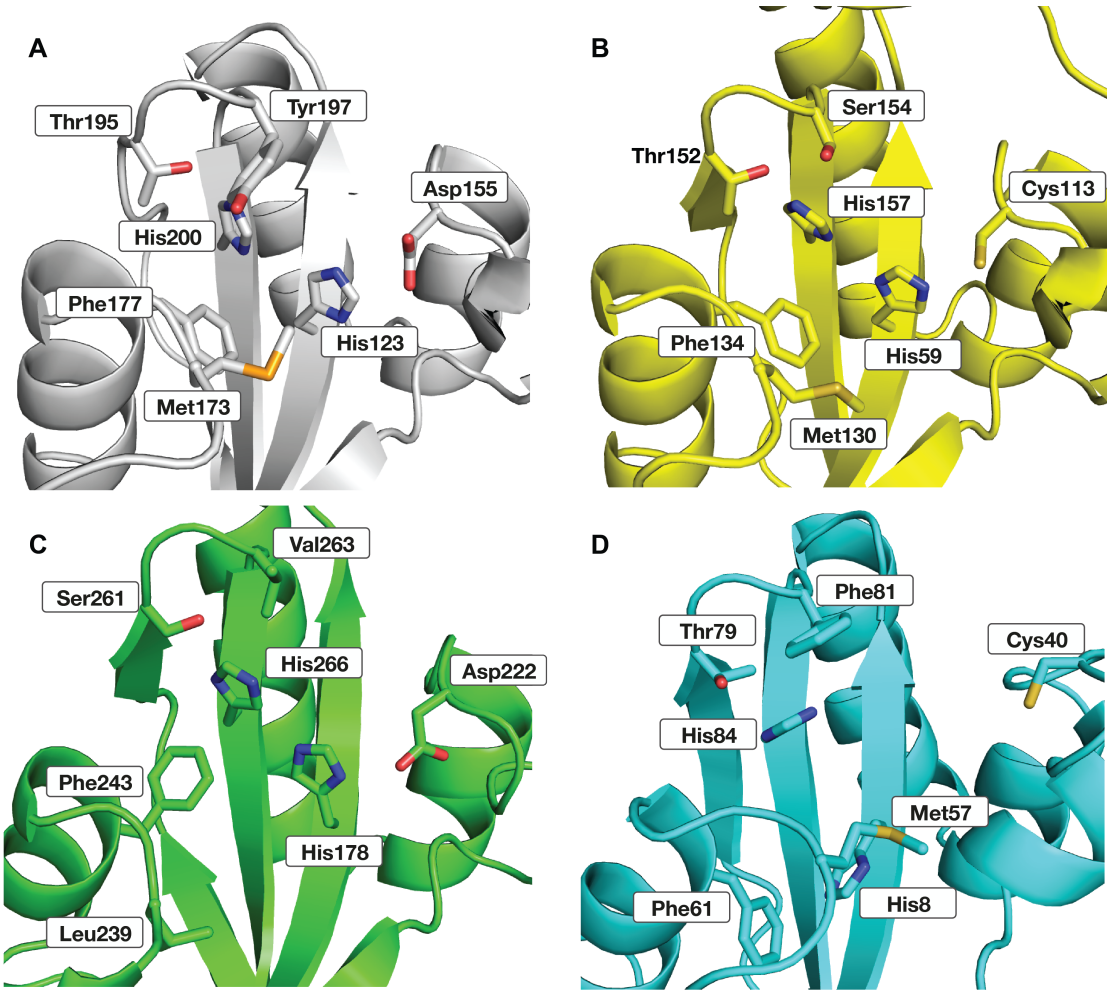
the NC domain, is almost exclusively α-helical, except for two short strands in the N-terminal region, which form an antiparallel β-sheet.

The PrsA parvulin domain consists of a four-stranded antiparallel β-sheet surrounded by four α-helices ($\beta\alpha_3\beta\alpha\beta_2$ topology) and superimposes well with other parvulin-type PPIase domains[159,164,167,287,308,319-324]. Structurally, it is most closely related to human Pin1[159], although PrsA is of prokaryotic origin and exhibits a different substrate specificity. The strongest differences locate to the S1–H1 loop (Figure 4.4C). In Pin1, this loop contains seventeen residues including a cluster of positively charged residues, which determine the specificity of Pin1 for phosphorylated Ser/Thr-Pro substrates. In PrsA, this loop is shortened to only a two-residue turn. The proline-binding pocket of PrsA is lined predominantly by hydrophobic residues. It also includes the highly conserved residues His123, Asp155, Thr195, and His200, which form an extended hydrogen bonding network[325] similar to the active sites of other parvulins (Figure 4.5).

PrsA crystallizes with two monomers in the crystallographic asymmetric unit, which superimpose with an overall root-mean-square deviation (rmsd) of 1.9 Å. The isolated domains are structurally very similar with rmsd values of 0.7 Å for the PPIase and 1.2 Å for the NC domains; the larger overall rmsd results from a slight hinge bending motion between the domains (Figure 4.4D). The dimer interface involves 47 mainly N-terminal residues, it extends over an area of 1830 $Å^2$ and is classified as stable[326]. In the PrsA dimer, the N-terminal ends are in close proximity, in agreement with dimer formation of the full-length protein anchored via its lipid-linked N-terminal cysteine to the membrane anchors in *B. subtilis* (Figure 4.4E).

4.4.4 Relation of NC domains to chaperone domains of SurA and trigger factor

In addition to their catalytic domains, foldases often contain domains for transient and dynamic interactions with their substrates. The N-terminal region of PrsA starts with a pair of short antiparallel β-strands (Val9-Thr13 and Asp16-Thr19) followed by five α-helices before entering the parvulin domain. The C-terminal region is arranged into two long α-helices that tightly interact with the N-terminal helices. Together with the parvulin domain, the N- and C-terminal helices form an extended crevice (Figure 4.4E).



E

PrsA	His123	Asp155	Met173	Phe177	Thr195	Tyr197	His200
Pin1	His59	Cys113	Met130	Phe134	Thr152	Ser154	His157
SurA-P2	His178	Asp222	Leu239	Phe243	Ser261	Val263	His266
Par10	His8	Cys40	Met57	Phe61	Thr79	Phe81	His84

Figure 4.5 Active site structure of PrsA.
A. PrsA in comparison to **B.** Pin1, **C.** SurA and **D.** Par10. Residues that contribute to the active site of the parvulin enzyme family are shown in stick representation and labeled. **E.** Comparison of the sequences that define the active site of PPIase domains.

Interestingly, despite a lack of sequence homology, the NC domain is structurally related to the chaperone domains of trigger factor, which assists the folding of newly synthesized proteins at the ribosome[327,328] and of SurA[322], a foldase for outer membrane proteins in Gram-negative bacteria. A related fold is also found in the less well-characterized leptospiral protein LIC12922. Figure 4.6 highlights the common structural elements of these four proteins. The 110 residues of PrsA (Met27–Leu115; Asp229–250) that build the central region of the NC domain superimpose well with SurA (Asp62–Ala161; Phe406–Ser427), trigger factor

(Ser280–Lys361) and LIC12922 (Ser89–Val184; Glu322–Leu351) with rmsd values of 3.3 Å, 3.6 Å and 3.1 Å[323], respectively. The “arms” and “body” structure described initially for trigger factor[329] is present in all four proteins (Figure 4.6A), but arm1 is significantly shortened in PrsA.

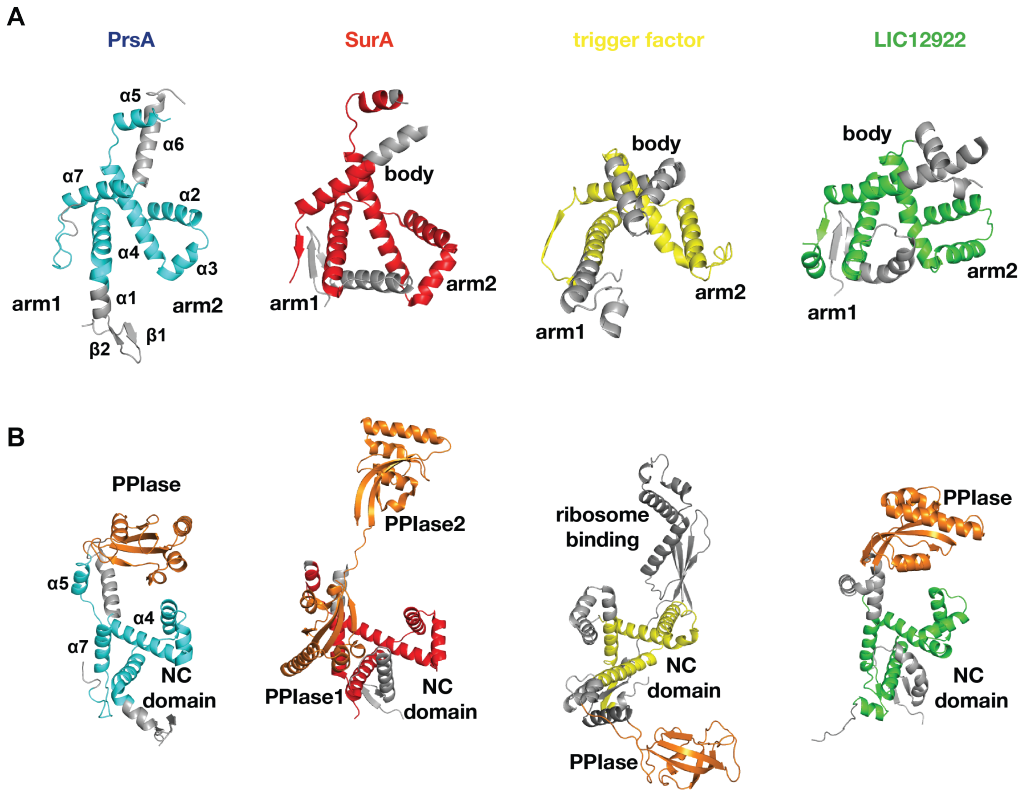


Figure 4.6 A common chaperone domain fold of the NC domain.
A. Cartoon representation of the PrsA NC domain (cyan) with the “body” and “arms” modules, in comparison with the chaperone domains of SurA (red) and trigger factor (yellow), both from *E. coli* and LIC12922 from *Leptospira interrogans* (green). Structurally non-conserved regions are represented in gray whereas conserved ones are highlighted. **B.** Orientation of the PPIase (parvulin) domains (orange), relative to the NC domain. The color code is the same as in (A). SurA contains two parvulin domains (ppiase1 and ppiase2). Trigger factor shows a FKBP-type PPIase and a ribosome binding domain (dark gray).

The linkers to the parvulin domain are relatively short (five amino acids) in PrsA, and therefore the prolyl isomerase site is located in close proximity to the NC domain, such that helix α2 of the NC domain is in close proximity (5 Å) to helix α3' of the parvulin domain (Figure 4.6B). The orientation of the two domains is similar in LIC12922, although the parvulin domain is tilted relative to its position in PrsA by about 25°, but differs from the organization of the well-characterized trigger factor and SurA chaperones. In trigger factor, an FKBP-type PPIase is inserted into the chaperone domain, and the linkers between the two domains are much longer. SurA consists of a chaperone domain and two parvulin-type

PPIase domains. In SurA, parvulin domain 1 is inactive and tightly associated with the chaperone (NC) domain, while the functionally active parvulin domain 2 is tethered to the NC domain also via a long, and presumably flexible linker[322] similar to the FKBP-type PPIase in trigger factor (Figure 4.6B).

4.4.5 Monomeric PrsA in solution does not prevent protein aggregation

Many periplasmic folding factors are known to have dual functions as PPIases and chaperones[330]. As the NC domain of PrsA shows structural similarity to chaperones we examined whether PrsA interacts with an unfolded protein. We performed citrate synthase (CS) aggregation and inactivation assays. GdmCl-unfolded CS aggregates spontaneously after dilution with refolding buffer, and this is accompanied by a strong increase in light scattering, unless aggregation is suppressed by a chaperone. PrsA was unable to inhibit the aggregation of CS (Figure 4.7), while 3.0 µM SlyD from *E. coli* completely abolished aggregation of CS[331]. Similar results have been previously obtained in a rhodanese aggregation assay[169]. Importantly, under the assay conditions, non-membrane tethered PrsA is monomeric and does not self-associate into its presumably native dimeric form. As a distinct chaperone activity has been observed for PrsA *in vivo*, these results strongly suggest that dimerization – in line with our structural data – may be a pre-requisite for chaperone function of PrsA, although we cannot rule out that PrsA may only be active towards a limited set of substrates different from CS.

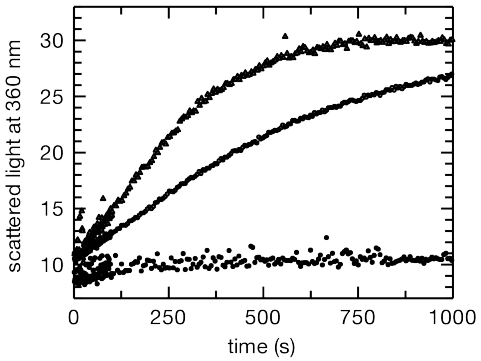


Figure 4.7 Monomeric PrsA in solution has no generic chaperone function.
Citrate synthase chaperone assay. Influence of PrsA on the aggregation of chemically denatured citrate synthase at 25 °C. Denatured citrate synthase was diluted to a final concentration of 0.15 µM (monomer) in 100 mM Tris-HCl (pH 8.0), 1 mM EDTA, 30 mM GdmCl, 50 mM NaCl, and 0.1 mM DTE. Light scattering at 360 nm was monitored in the absence of enzyme (○) and in the presence of 3.0 µM SlyD* (●), in the presence of 3.0 µM PrsA (Δ).

4.4.6 Binding to presequence involves residues from parvulin and NC domains

In Gram-positive organisms, exported proteins commonly contain a propeptide sequence, which follows the export signal sequence and increases export efficiency[285]. The AmyE propeptide from the alpha-amylase of *B. subtilis* is eleven amino acids long (ETANKSNELTA) and contains no proline residues. When attached to other proteins, it improves their export[285,332]. If PrsA is overexpressed in addition to AmyE attachment, the protein export is further increased[285], suggesting functional interplay and possibly a direct interaction of PrsA and the AmyE propeptide.

To examine how PrsA interacts with the AmyE propeptide, we recorded two-dimensional [^{15}N,^{1}H]-TROSY spectra of ^{15}N-labeled PrsA in the presence of the AmyE-peptide at increasing concentrations. The [^{15}N,^{1}H]-TROSY cross peaks of [^{1}H,^{15}N]-PrsA were well resolved and enabled a detailed analysis of the chemical shift changes upon addition of the AmyE-propeptide (Figure 4.8A). The resonances of the isolated parvulin domain of PrsA (residues 116–206,[287]) closely match corresponding resonances of full length PrsA. This indicates that most amide moieties of the parvulin domain in isolation and in full length PrsA are in similar chemical environments and allows a direct transfer of sequence-specific resonance assignments for the analysis of chemical shift changes induced by binding of the AmyE propeptide to full length PrsA.

Addition of 5.0 mM of the AmyE propeptide to 0.5 mM PrsA (final ligand to protein ratio 10:1) led to chemical shifts larger than 0.05 ppm for the amide moieties of 18 residues in the parvulin and 37 residues in the NC domain (Figure 4.8B). In all cases, the resonances changed their positions continuously during the titration (Figure 4.8C, D), indicating that exchange kinetics between bound and unbound form are occurring in the fast exchange limit of the NMR chemical shift timescale (> 1000 s^{-1}). The residues in the PPIase domain with the largest chemical shift changes are located around the active site (Figure 4.8E). From the concentration dependence of the chemical shift change we derived dissociation constants for the interaction between the propeptide and individual residues in the PPIase or NC domain (Table 4.2; Table 4.3; Figure 4.8C, D). With the caveat that substrate saturation was not reached for all residues in this experiment, the data are well suited to establish an overall estimate of the binding constant of the

substrate peptide to the PPIase (in the range of 1.5 to 4.5 mM) and NC domain (in the range of 1.2 mM to 4.9 mM).

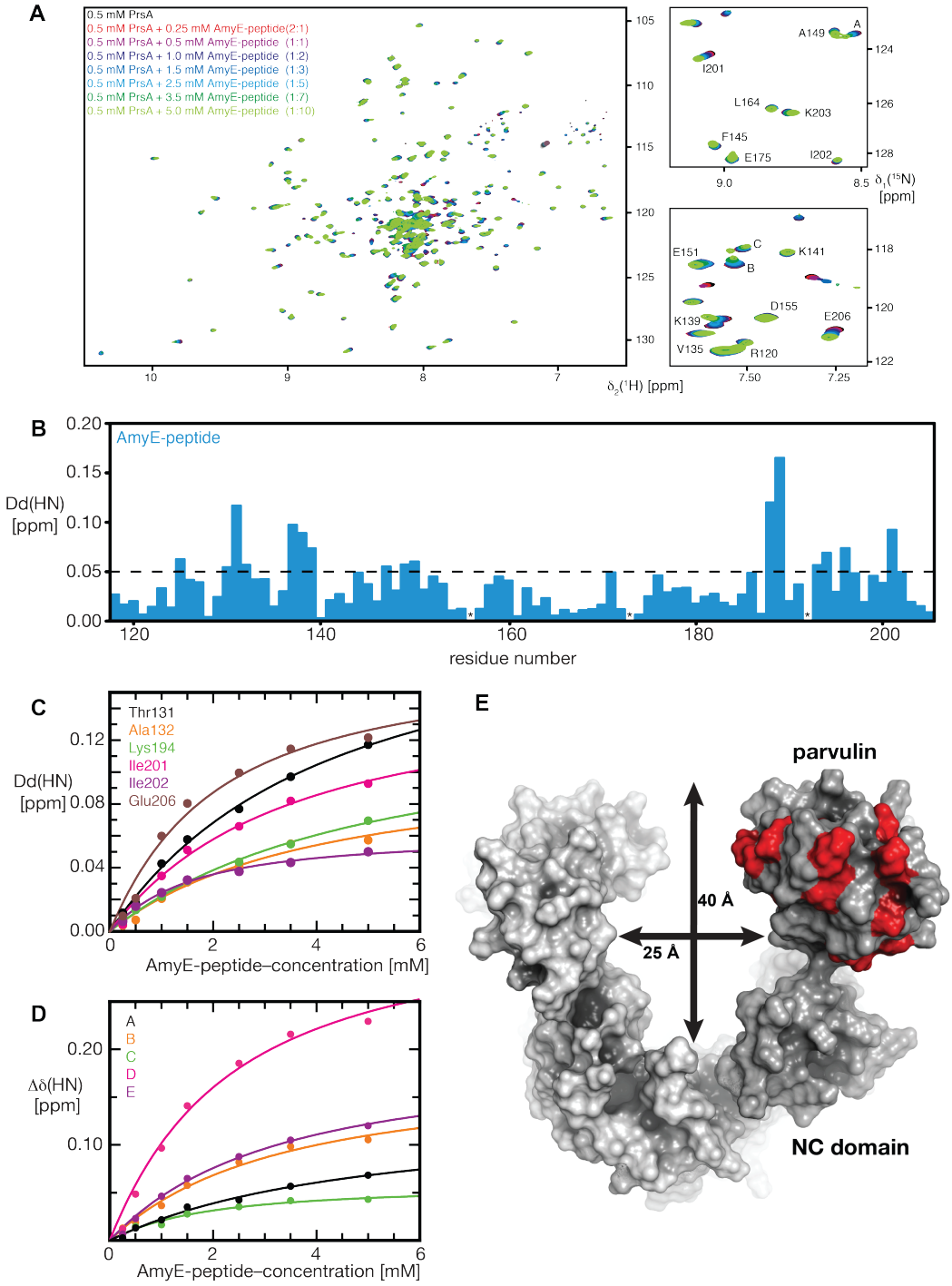


Figure 4.8 Mapping of PrsA residues involved in AmyE-propeptide binding.
A. Left hand panel: [^{15}N,^{1}H]-TROSY spectra of 0.5 mM [U-2H,^{15}N]-PrsA in 25 mM MES pH 6.5 with increasing AmyE-peptide concentrations, as indicated. The protein to ligand ration is given in parentheses. Right hand panels: Enlargement of two regions with the assignment derived from the NMR-structure of the PPIase-domain alone ([287]; BMRB 6601) as well as peaks belonging to the NC-domain, that were analyzed (Table S1). All spectra were measured at 25°C with eight scans and 1024 x 200 complex points. **B.** NMR peak shift analysis of AmyE-peptide to PrsA titration. Combined chemical shift changes of the amide moiety upon titration with a 10-fold excess of AmyE-peptide (blue)

vs. the residue number. The broken line indicates a significance level of 0.05 ppm. '*' denotes unassigned residues as well as Pro192. **C.** Backbone chemical shift perturbations for residues located in the parvulin domain of 0.5 mM PrsA upon titrating increasing AmyE-peptide. **D.** Backbone chemical shift perturbations for residues located in the NC domain of 0.5 mM PrsA upon titrating increasing AmyE-peptide. The solid lines in B, C represent nonlinear least square best fits of the normalized changes of the 1H and ^{15}N chemical shifts to all of the titration data simultaneously, using a bimolecular equilibrium binding model. **E.** Surface representation of the PrsA dimer (monomer A in dark gray and monomer B in light gray). PrsA residues of the parvulin domain involved in AmyE-propeptide binding identified by NMR are shown in red. The dimensions of the PrsA cavity are indicated.

These dissociation constants are identical within the experimental precision. Overall, the NMR experiments thus demonstrate that PrsA directly interacts with the AmyE propeptide and that both the NC and the parvulin domain contribute to the interaction, even though the propeptide does not contain proline residues.

Table 4.2 Interaction of the AmyE-peptide with PrsA. Dissociation constants for the interaction of PrsA with the AmyE-peptide for selected PrsA residues of the parvulin domain, as determined by NMR spectroscopy titrations.

Residue	AmyE K_D [mM]
Thr131	3.9 ± 0.5
Ala132	3.7 ± 1.3
Lys194	4.8 ± 0.7
Ile201	3.1 ± 0.7
Ile202	1.2 ± 0.2
Glu206	1.7 ± 0.5
Median	**3.4 ± 1.6**

Table 4.3 Interaction of the AmyE-peptide with the NC domain of PrsA. Dissociation constants for the interaction of PrsA with AmyE-peptide for selected PrsA residues of the NC-domain, as determined by NMR spectroscopy titrations.

Residue	^{1}H [ppm]	^{15}N [ppm]	PrsA K_D [mM]
A	8.52	123.4	4.5 ± 1.2
B	7.52	118.5	2.9 ± 0.7
C	7.50	118.0	1.5 ± 0.4
D	8.00	117.8	2.0 ± 0.7
E	7.44	112.3	2.9 ± 0.5
Median	**—**	**—**	**2.8 ± 1.1**

4.5 Discussion

In Gram-positive bacteria, most secreted proteins are translocated in an unfolded state into the space between the cell membrane and the cell wall[333,334]. This compartment is characterized by high concentrations of cations bound to teichoic acid, a high negative charge density, and a low pH[335], which together form a challenging environment for protein folding. PrsA-like proteins are key folding factors to support post-membrane protein maturation and secretion in Gram-positive bacteria[336]. The crystal structure of PrsA demonstrates that PrsA folds into two subdomains, the chaperone-like helical NC domain and the PPIase domain, which is inserted into the NC domain. PrsA shares this domain architecture with folding factors from Gram-negative organism such as trigger factor or SurA[322], despite a lack of sequence homology in the respective N- and C-terminal regions. In PrsA, PPIase activity is independent of the presence of the NC domain, as demonstrated by the fact that the isolated, monomeric parvulin domain resembles the catalytic properties of the full-length protein[169,322]. It should also be noted that other chaperones such as prefoldin, Skp, or the mitochondrial small Tim proteins also use helical protrusions ("tentacles") to interact with folding proteins[330,337,338].

Despite the structural homology of the NC domain to the chaperone domains of other folding factors, non-membrane tethered PrsA is not active in common chaperone assays (Figure 4.7). So is the NC domain a true chaperone relevant for PrsA function? *In vivo* experiments have clearly demonstrated that the isolated NC domain alone is sufficient for promoting the secretion of various substrates and even rescues a PrsA deletion knockout[169,174]; still all domains of a PrsA homologue in *L. monocytogenes* (PrsA2) are required for full virulence[174]. Together, these results indicate, that the PPIase activity of PrsA is relevant for a subset of substrates, but that the most essential functions of PrsA are mediated by the NC domain. The most likely reason for the absence of chaperone activity of PrsA in solution assays is a requirement on dimerization, although other reasons, such as mismatch between test substrates and PrsA specificity cannot be completely ruled out at this point. We demonstrate that PrsA lacking its membrane anchor is monomeric, but has an inherent tendency for dimer formation at high protein concentration.

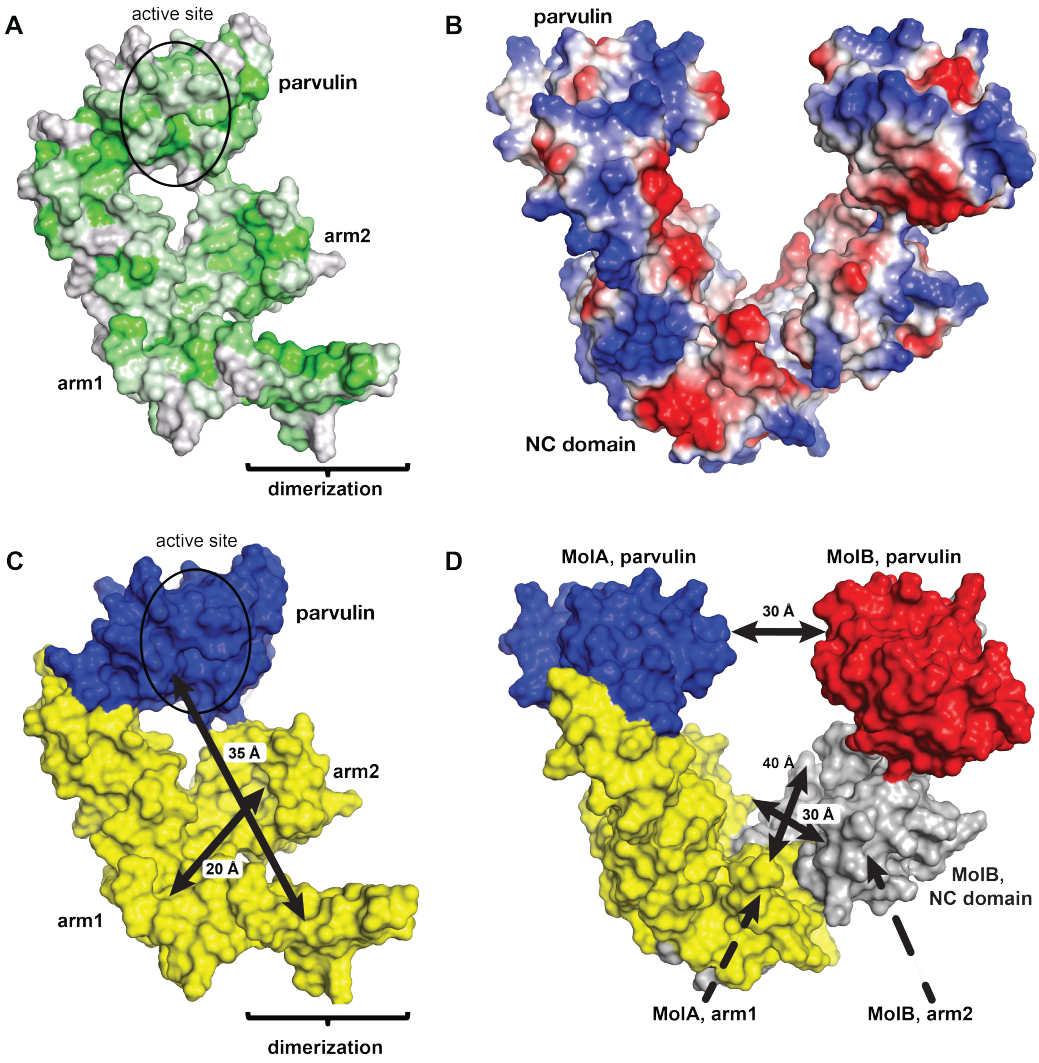


Figure 4.9 Surface properties and dimensions of PrsA.
A. Surface representation of PrsA color coded ranging from hydrophobic (green) to hydrophilic (gray) according to the normalized consensus hydrophobicity scale of the exposed residues[339]. **B.** Surface representation of PrsA color-coded ranging from negative charged (red) to positive charged (blue). **C.** The crevice of the PrsA dimer can accommodate domain-sized polypeptides. Monomer structure of PrsA; The parvulin domain and the NC domain are colored blue and yellow, respectively. The PPIase active site is indicated as ellipse. The arm1-arm2 distance (20 Å) as well as the distance from the parvulin domain to the opposite site of the NC domain (35 Å) within one PrsA molecule are given. **D.** Representation of PrsA dimer with approximate dimensions of the cavity indicated; PrsA molecule A, is colored as in (C). For PrsA molecule B, the parvulin domain and the NC domain are colored red and gray, respectively. The intermolecular distance of arm1, arm2 and parvulin regions are 28 Å, 40 Å and 29 Å, respectively.

It crystallizes in a dimeric arrangement, which is fully compatible with membrane tethering and employs exclusively the NC domain for dimerization. *In vivo*, lipid anchoring dramatically favors dimer formation, and in fact, PrsA2 of *L. monocytogenes* and PrsA of *B. subtilis* have been reported to form dimers at the cell membrane[173,174].

In the dimeric assembly, PrsA forms a bowl-like crevice (Figure 4.4E) between the NC domains of two monomers. The inside surface of the crevice is enriched in hydrophobic residues (Figure 4.9A), whereas the outside contains many charged amino acids (Figure 4.9B). Analogous hydrophobicity profiles are observed for the NC domains of related *bona fide* chaperones, e.g. trigger factor[329], suggesting that the inside of the crevice could provide a chaperone function *in vivo*. Enlarging the interaction surface for unfolded peptides in the dimer might be critical to permit productive substrate interaction by avidity of multiple, low affinity interactions. With dimension of about 20 x 30 x 35 Å (Figure 4.9C, D), the bowl-shaped crevice in the PrsA dimer could completely accommodate polypeptides of up to 20 kDa, but the observed flexibility of the NC domains (Figure 4.4D) and the attached parvulin domain may even permit adaptation to larger substrates. The only other chaperone with PPIase/NC domain architecture, for which dimerization and the formation of an enlarged crevice has been observed is Peb4 from *Campylobacter jejuni*[167]. However, its NC region is more distantly related to SurA/trigger factor and dimerization under unusual domain-swapping is required for completion of the NC-domain fold. Many Gram-positive bacteria contain only a single PrsA-like protein. Others, such as *B. cereus*, contain several PrsA homologs[336]. For *L. monocytogenes* and *B. anthracis* it was shown that PrsA homologs have non-overlapping functions and cannot substitute each other[176,177]. Interestingly, the parvulin domains, but not the NC chaperone domains of PrsA homologues are highly conserved (Figure 4.10). The high conservation in the PPIase domain is necessary to preserve the enzymatic function; the variations in the NC-domain may reflect differing substrate specificities.

PrsA substrates exert functions in cell wall metabolism (penicillin binding proteins), swimming motility (flagellin), oligopeptide transport (OppA), and membrane bioenergetics (quinol oxidase), act as virulence factors or as enzymes such as α-amylase or subtilisin[336,340]. The molecular weight of substrates ranges from 20 to 80 kDa, and thus includes proteins, which are clearly too large to be accommodated in the PrsA crevice. The lateral openings of the crevice might serve to provide the possibility for interactions of multiple PrsA molecules with one substrate polypeptide, as found for trigger factor, where e.g. three molecules bind to a 45 kDa protein[329].

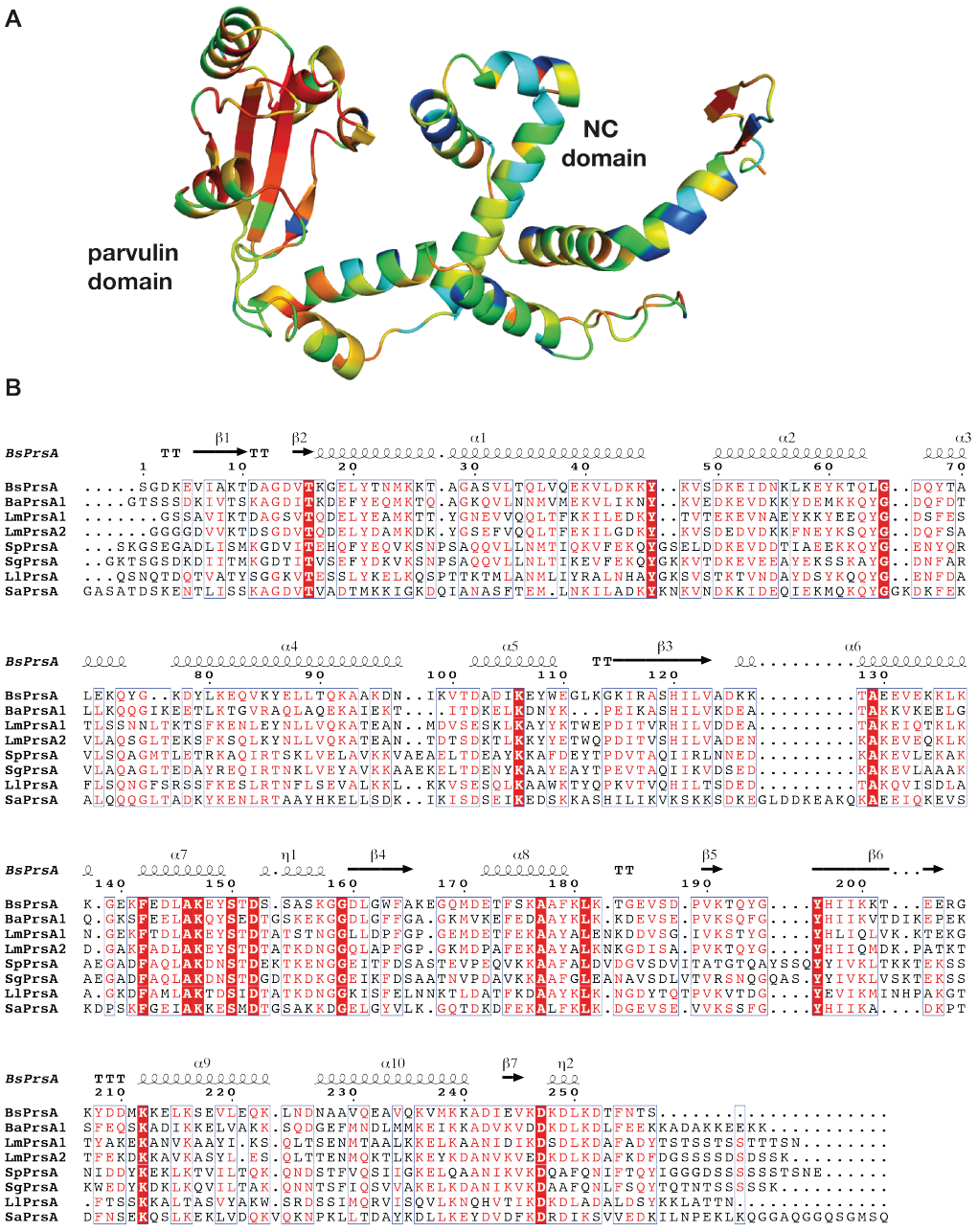


Figure 4.10 Sequence conservation in the PrsA family.
A. Consurf representation[341] of sequence conservation within the PrsA family; residues that are highly conserved are shown in red, sequences with lower identity are shown in yellow, green and blue. **B.** Multiple sequence alignment to analyze the sequence and secondary structure conservation of PrsA. Highly conserved residues are red (>70% conservation) or white in red boxes (100% conservation). The secondary structure of B. subtilis PrsA is shown on top of the protein sequence. Sequences of representative PrsA proteins were retrieved from the UniProt database[342] and aligned using MULTALIN[343]. The final figure was generated using the ESPript server[344]. Species abbreviations and UniProt accession numbers are BsPrsA, *Bacillus subtilis* (P24327); BaPrsA1, *Bacillus anthracis* (Q81U45); LmPrsA1 *Listeria monocytogenes* (Q71ZM6); LmPrsA2, *Listeria monocytogenes* (Q71XE6); SpPrsA, *Streptococcus pneumoniae* (B2IPD4); SgPrsA, *Streptococcus gordonii* (A8AYJ0); LlPrsA, *Lactococcus lactis* (P0C2B5); SaPrsA, *Staphylococcus aureus* (A6QI23).

We show that PrsA directly interacts with substrate propeptides and analyze the binding of PrsA to a prototypic eleven residues peptide. The NMR experiments indicate that PrsA uses both the NC chaperone and the parvulin domain for transient and dynamic substrate interactions.

To our knowledge, PrsA is the first authentic member of the SurA/trigger factor class of foldases, which has been structurally characterized as a symmetric dimer and may require dimer formation for exerting chaperone activity. Our current structural and functional characterization of PrsA provides a basis for future in-depth studies on the role of self-association e.g. by targeted trapping of dimeric states for solution studies. The structural information will help to identify PrsA regions involved in polypeptide recognition, and to understand its role in the maturation of larger proteins. This paves the way for new strategies to optimize protein export in biotechnological applications and may yield opportunities for developing small molecule inhibitors of extracellular protein folding as antibacterial agents against Gram-positive pathogens.

4.6 Supplemental Data

Supplemental Table 4.1 Statistics on diffraction data and structure refinement of PrsA

	PrsA
Wavelength (Å)	1.36246
Space group	C 2 2 2_1
Unit cell	69.29 87.68 234.1 90 90 90
Resolution (Å)	117-2.63 (2.79-2.63) *
Total reflections	144433
Unique reflections	21428
Multiplicity	6.7 (5.4)
Completeness (%)	98.8 (90.2)
Mean I/sigma(I)	19.7 (1.6)
Wilson B-factor	90.7
R-merge	0.054 (1.097)
$CC_{1/2}$	0.998 (0.65)
R-work	0.2160 (0.4199)
R-free	0.2470 (0.4284)
Number of atoms	8281
macromolecules	4079
water	37
Protein residues	513
RMS(bonds)	0.009
RMS(angles)	0.95
Ramachandran favored (%)	97
Ramachandran outliers (%)	0.2
Clashscore	1.7
Average B-factor	109.1
macromolecules	109.3
ligands	
solvent	89.9

5 Discussion & Outlook

5.1 Summary of results

The bacterial extracellular foldase PrsA in its membrane-bound form is a dimeric protein responsible for folding of secreted proteins in Gram-positive bacteria. The PPIase subdomain is highly conserved, while the NC exhibits higher variability. The NC domain shares a common fold with chaperone domains of cytosolic and periplasmic folding helpers, such as trigger factor and SurA. The dimeric, membrane-anchored PrsA creates a bowl-like crevice that is enriched with hydrophobic residues on the inside, where both the PPIase and the NC domain interact with the substrate.

PrsA is the first reported foldase of the SurA/trigger factor family that dimerizes in its active form, and the presented results may guide future optimization of protein export in biotechnological applications. Research groups will continue to work on this and related systems in order to reveal functional principles of chaperone action, e.g. with respect to surface properties or shape parameters. A further area of interest is the role of protein-protein interaction in the biogenesis pathways of secreted and outer-membrane proteins.

The core focus area of this thesis was the structural and functional characterization of biotin-dependent carboxylase multienzymes. We provide a structural analysis of yeast ACC, which reveals a detailed mechanism for phosphorylation-based short-term regulation. In detail, we obtained a crystal structure of the unique ACC central domain of *S. cerevisae* at 3.0 Å resolution, along with several intermediate- and low-resolution structures of larger fragments up to near-full length constructs. In combination with EM and SAXS analyses, these results reveal an overall elongated architecture with the CD acting as a flexible hinge between the rigid CT dimer and the BC. ACC inhibition is mediated by a regulatory loop, which is phosphorylated at a strictly conserved serine residue. In its phosphorylated state, this loop wedges into a crevice between two domains of the CD and restricts conformation freedom of interdomain orientation. A model is proposed where the removal of the phosphate group would reduce the affinity of the loop for interdomain binding, resulting in

increased flexibility of CD that is required to adopt a competent state for catalysis.

The hybrid structure of YCC from *D. radiodurans* provides novel insights into active site architecture, overall assembly and dynamic architecture of acyl-CoA carboxylase multienzymes. EM and SAXS experiments elucidate the flexible tethering and mobility of the BC domains relative to the rigid CT core, which are key characteristic and unique functional features of singe-chain bacterial carboxylase multienzymes.

The following sections will focus on discussing the structure, function and regulation of multienzyme carboxylases in the context of general multienzyme architecture. Prospects of structural studies on multienzymes for combinatorial biosynthesis and synthetic biology will be outlined.

5.2 A compact structure of active full-length yeast ACC

Most recently, Wei & Tong reported a structure of full-length yeast ACC (flACC) in biotinylated and unbiotinylated form along with several structures of CD fragments[345]. In flACC, the ACC dimer obeys two-fold symmetry and assembles in a triangular architecture, with dimeric BC domains and a large channel through the center of the compact assembly (Figure 5.1). The CT dimer and the C-terminal half of CD form one edge of the triangle, while the N-terminal half of CD constitutes for the other edges, with the BC dimer and the CD_{C1} serving as the corners. The mobile BCCP domain was resolved bound to the active site of the CT. Despite the modest resolution and the fully independent structure determination, the novel CD structures reported by Wei & Tong agree very well with the CD structure presented in chapter 2 at the level of individual subdomains.

However, the overall architecture of the CD and the entire ACC depicted in our work and by Wei & Tong differs dramatically. The transition from an elongated shape that is presented in this thesis towards a compact triangular shape is based on an intricate interplay of several hinge-bending motions in the CD, which are discussed in detail below.

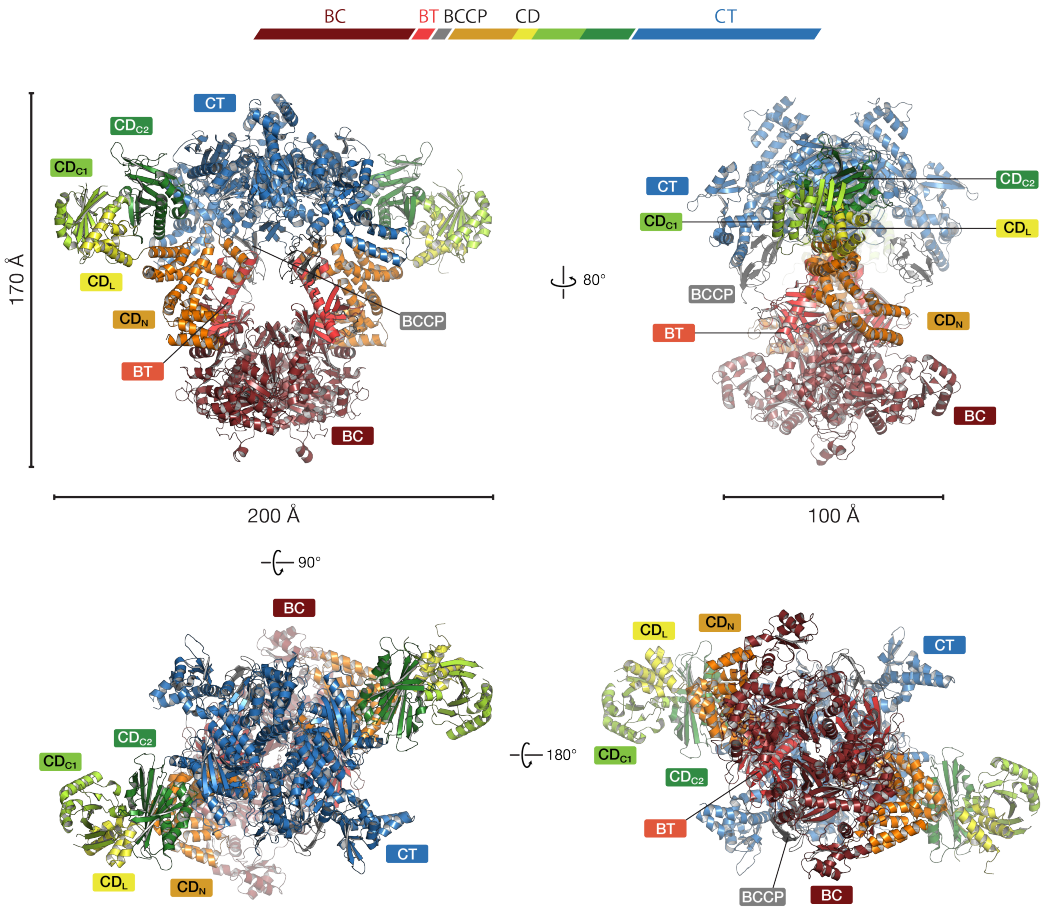


Figure 5.1 Compact crystal structure of active yeast flACC.
The crystal structure of yeast flACC expressed in *E. coli* exhibits a triangular architecture with dimeric BC domains. The domains of both protomers are colored according to the sequence scheme. Dimensions are indicated and domains of one protomer are labeled.

The largest conformational change is located at the CD_{C2}-CT hinge (Figure 5.2A,B). The angle between CT and CD_{C2}, tethered via a connecting helix to the N-terminus of the CT domain, differs between the two models by about 120°. Propagating this transition all the way to the N-terminal BC domain would already results in a displacement of BC by >200 Å. The conserved RxxGxN on the surface of the CT N-lobe, which interacts with the CD_{C2} domain in the elongated ACC model, is solvent exposed without further interactions in the compact form of flACC.

A second hinge is identified between the CD_{C1} and CD_{C2} domains (Figure 5.2C). In this thesis, this hinge was described as blocked, allowing only limited tilting and rotation, due to the wedging of the phosphorylated regulatory loop into the CD_{C1}/CD_{C2} interface. Since the protein was expressed in *E. coli* for flACC structure determination, this loop is not phosphorylated in flACC. Consistent with the regulatory mechanism proposed in this work, the non-phosphorylated

regulatory loop remains disordered in flACC. Concomitantly, the elongated and compact models of ACC differ by a considerable rotation of about 30° around the CD_{C1}/CD_{C2} hinge. This rotation leads to a displacement of the CD_N anchor point by 21 Å.

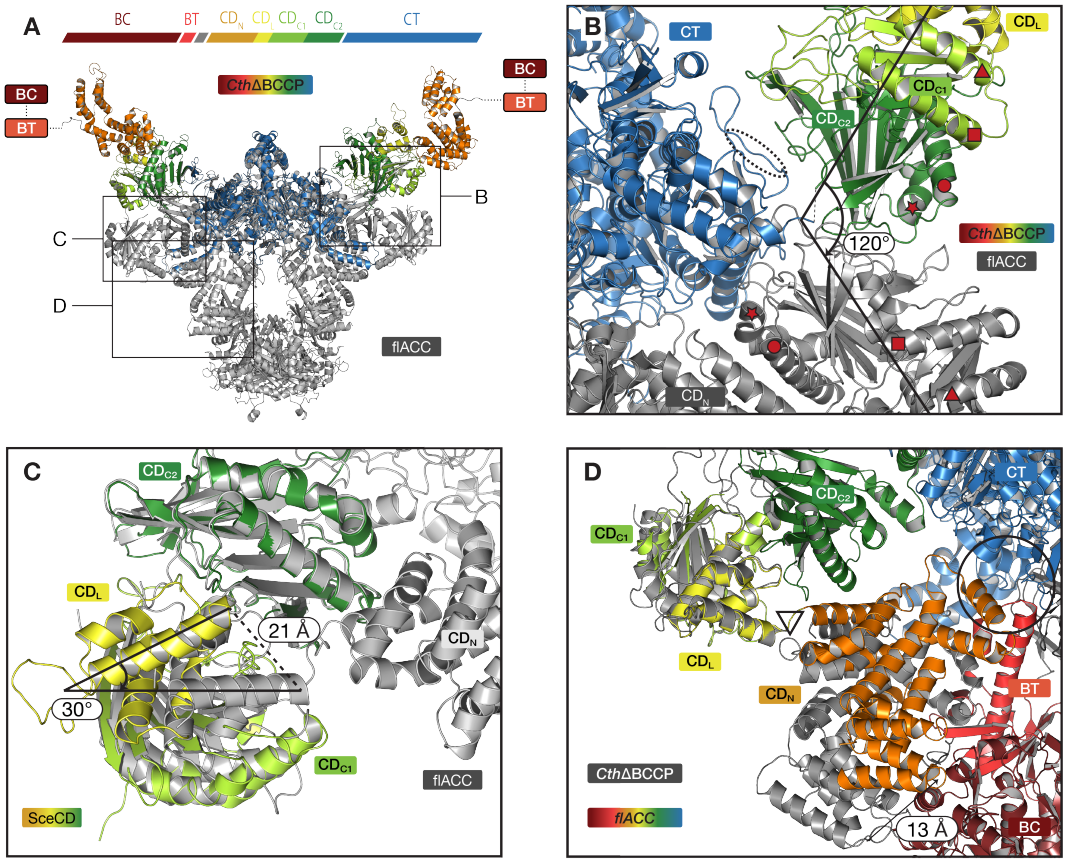


Figure 5.2 Conformational freedom governed by three main flexible hinges.
A. CT based overlay of *Cth*ΔBCCP (colored according to sequence scheme) and flACC (gray, same view as in Figure 5.1), displaying strikingly different orientation of the CD. Squares mark the regions that are shown in detail, viewing angles may diverge from A. **B.** CT based overlay (*Cth*ΔBCCP in color, flACC in gray) demonstrating the extensive hinge-motion (120°) occurring at the connecting helix. Position of the conserved RxxGxN is marked by a dashed ellipse. Corresponding structural elements in CD_{C1} and CD_{C2} are marked with red symbols (star, triangle, square, circle) for better orientation. CT of flACC was omitted for clarity and all visible domains are labeled. **C.** Superposition based on CD_{C2} (*Sce*CD in color, flACC in gray) demonstrating the 30° angular movement in this hinge between phosphorylated and non-phosphorylated ACC. The anchor point to CD_N is displaced by 21 A. **D.** Superposition based on CD_{C1} (*Cth*ΔBCCP in gray, flACC in color). The displayed *Cth*ΔBCCP is the instance that was observed to be closest to flACC. Still, an additional translation by 13 Å of CD_N at the end distal to the interaction between BT, CD_N and CT (marked with a circle) is observed. All visible domains are labeled according to the sequence scheme and the hinge is denoted with a black triangle.

When comparing all individual protomer and fragment structures in their study, Wei & Tong also identify the CD_N/CD_1 connection as a highly flexible hinge, in

good agreement with our observations (Figure 5.2D). In the flACC structure, however, the hinge is fixated by interactions of CD_N with BT and CT in a particular conformation that was not observed in our work and results in a 13 Å translational offset of the distal part of CD_N between the elongated and compact ACC models.

The only bona-fide regulatory phosphorylation site of fungal ACC in the regulatory loop is directly participating in CD_{C1}/CD_{C2} domain interactions and thus influences the hinge conformation at the CD_{C1}/CD_{C2} hinge. Modeling the CD_{C1}/CD_{C2} hinge conformation observed here in the phosphorylated *Sce*CD onto the compact flACC structure demonstrates that phosphorylation and subsequent wedging of the regulatory loop prevents the formation of a compact flACC structure with catalytically competent dimeric BC: When keeping all other hinges fixed as in flACC, the CD as well as the BC would severely clash with other parts of the protein, and the two BC domains would be separated by >150 Å. The catalytically competent conformation could also not be restored by applying any combination of hinge-bending motions observed by Wei & Tong or in this work.

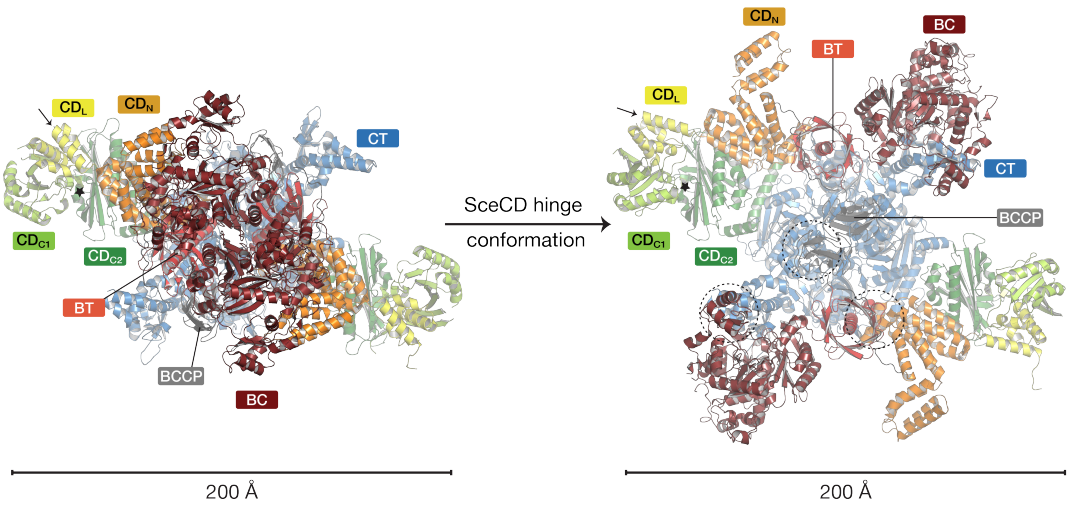


Figure 5.3 *Sce*CD hinge conformation distorts flACC architecture.
On the left, the same view as in Figure 5.1, all domains of one protomer are labeled and the scale is indicated. On the right, the architecture after applying the CD_{C1}-CD_{C2} hinge conformation from *Sce*CD. A large distortion of the architecture can be observed leading to severe clashes (circles) and to the BC domains being separated by >150 Å. The arrows point to helix described in Figure 5.2C and a black star marks the CD_{C1}-CD_{C2} hinge. The same domains as on the left side are labeled.

Further comparative mutational and structural studies will be required to analyze whether the binding of the phosphorylated regulatory loop to CD_{C1}/CD_{C2} has

long-range conformational effects, plausibly mediated via allostery by contacts of CD_{C1}/CD_{C2} to other domains. Amongst all ACC fragment structures, the orientation of CD_N in our isolated yeast CD structure is unique, and differs from other fragments and flACC by a 180° rotation relative to CD_L/CD_{C1}. Modeling the rotated orientation of CD_N onto the flACC structure yields a plausible structure without steric clashes. It thus currently remains open, whether the rotated orientation of CD_N also occurs in the context of intact yeast, eventually with lower occupancy. It should be noted that Wei & Tong observed in isolated structures of phosphorylated CD domain fragments (comprising BT, BCCP and CD) the same conformation of the CD_{C1}/CD_{C2} hinge that we observe in our phosphorylated yeast CD structure, but with a disordered regulatory loop. This clearly indicates that regulation occurs by conformational selection, where binding of the phosphorylated loop considerably increases the population of a pre-existing extended ACC conformational state.

Altogether, the comparative structural analysis clearly establishes a unique mechanism of fungal ACC based phosphorylation-mediated conformational trapping of the characteristic CD domain.

5.3 Implications for the regulation of eukaryotic ACC

Yeast ACC in its active form adopts a compact, triangular structure. Mutational analysis of the BC dimer interface demonstrates that BC dimerization in the compact flACC is relevant for activity[345]. A functional role of BC dimerization is also indicated by a lack of activity of monomeric eukaryotic ACC BC in solution. As described in detail above, the CD_{C1}/CD_{C2} hinge conformation observed in the structure of phosphorylated CD is not compatible with BC dimerization in the holo-enzyme, explaining why the phosphorylated form is in an elongated shape with flexible BC-CD arms, relative to the dimeric CT, as confirmed by EM. This establishes a direct link between phosphorylation of the conserved Ser1157, the overall conformation and the enzymatic activity and thus reveals a conformational trapping mechanism of phosphorylation control of yeast ACC[80,131,132].

Preliminary activity assays (unpublished), based on incorporation of radioactive carbon into stable malonyl-CoA[193], support the role of the Ser1157 phosphorylation site. An increase in activity is observed after treatment of

phosphorylated ACC with λ protein phosphatase, which decreases levels of Ser1157 by at least 60-fold based on MS analysis. In addition, a similar increase in activity is observed for a non-phosphorylatable S1157A mutant. Mutational analysis of the arginine residues involved in phosphor-Ser1157 binding confirms their involvement in yeast ACC regulation: Arginine to alanine mutations of the coordinating Arg1173 and Arg1260 restore activity of phosphorylated yeast ACC. These preliminary experiments are currently being extended to obtain a valid functional mapping of the control mechanisms in yeast ACC.

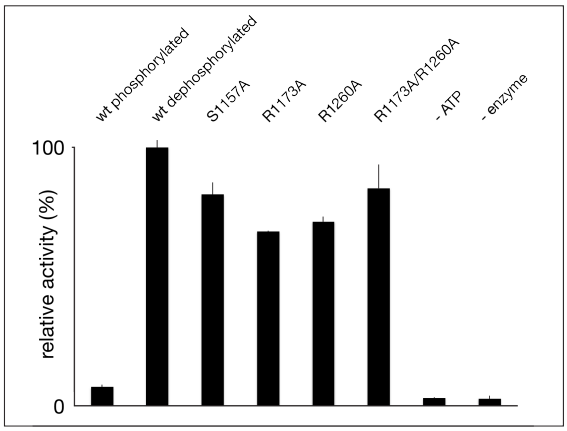


Figure 5.4 Dephosphorylation of Ser1157 activates *S. cerevisae* ACC.
Activity assay based on incorporation of radioactive ^{13}C into acid-stable products. Activity of wild-type (wt) ACC increases drastically upon dephosphorylation. A non-phosphorylatable mutant (S1157A) and arginine to alanine mutants (R1173A, R1260A, R1173A/R1260A), incapable of phospho-serine coordination, display similar levels of activity as observed for dephosphorylated wild-type ACC. All values normalized to the activity of dephosphorylated wild-type ACC. Error bars are based on triplicate measurements. – ATP and – enzyme mark controls, where either ATP or ACC was omitted from the reaction.

Human ACC is also regulated by phosphorylation and our crystal structure of human CD clearly resembles instances of yeast CD observed in various yeast ACC fragments or flACC. However, based on biochemical analysis, the mechanism of regulation diverges from those of the homologous yeast ACC. Reversible phosphorylation of human ACC mediated by AMPK and PKA occurs at Ser80 and Ser1201[127]. Ser1201 is located in the equivalent of the yeast ACC regulatory loop and it would be tempting to speculate that inhibition via Ser1201-phosphorylation works similarly as Ser1157 phosphorylation in yeast ACC, particularly because the phosphate-coordinating arginines R1173 and R1260 of yeast ACC are conserved in human ACC. Ser1201. However, Ser1201 does not align to Ser1157 in yeast and is only moderately conserved in higher eukaryotes.

It is also not phosphorylated to high-levels and consequently disordered in our human CD domain. Earlier biochemical analysis revealed that phosphorylation of Ser1201 alone has no significant role in control of human ACC.

Instead, regulatory inhibition is mainly driven by phosphorylation of Ser80[126], which is not conserved in yeast ACC. Ser80, conserved amongst all higher eukaryotes, resides in a flexible N-terminal extension preceding the BC domain. Upon phosphorylation, Ser80 binds into the dimer interface of BC, preventing dimerization. Interestingly, the antifungal and antibiotic compound Soraphen A binds into the same pocket in yeast and human BC, and mimicks the inhibitory effect of phosphorylated Ser80 binding.

Thus, while human and yeast ACC are controlled via regulating the dimerization of the BC domain, they employ to completely distinct strategies: Direct interference with interface formation vs. allosteric control. Principal components of both regulatory systems, the Soraphen A–binding site and the regulatory loop, appear to be present in both, human and yeast ACC. This may indicate that both systems have evolved from a common precursor, which used both control pathways and opens the possibility that some current ACC variants may still do so.

Human ACC in its most active form assembles into large symmetric linear polymers (Figure 5.5A)[93,187]. A direct link between phosphorylation and polymerization is not established yet, but could eventually explain why another mechanism of regulation is employed by human ACC compared to yeast ACC. The mechanism of locking the CD conformation, as seen in phosphorylated yeast ACC, may be incompatible with polymerization and might thus have lost its role in controlling BC dimerization.

There are no data available on how human ACC polymers are assembled at a structural level. Conceptually distinct modes of interaction could be envisioned based on ACC structure (Figure 5.5B): (i) Filaments could form via CD-CD contacts under structural rearrangement of the CD and eventually the entire assembly. (ii) Filament formation may result from dimerization of BC domains from different dimeric ACC molecules. This would lead to a sandwiched assembly of CT dimers at the bottom and on top, with four BCs between them, where one BC belonging to the bottom dimer is contacting a BC of the upper dimer. This may result, as in the first model, in modulations of CD geometry and altered active site distances, ultimately leading to increased activity. To form a

filament in this model, however, CT dimers have to interact and form stable dimers of dimers with the next CT dimer, which has not been reported so far. (iii) Filaments could form by interactions of BC alone, similar to the second assembly mode, outlined above, but without further contributions from the CT. (iv) They could also form by combined interactions from BC and CD or (v) CD and CT, respectively. While some of these qualitative models cannot explain the increased activity, others seem to be incompatible with the EM micrographs. The assembly mode of these filaments remains one of the big mysteries in the field and will be addressed in future research.

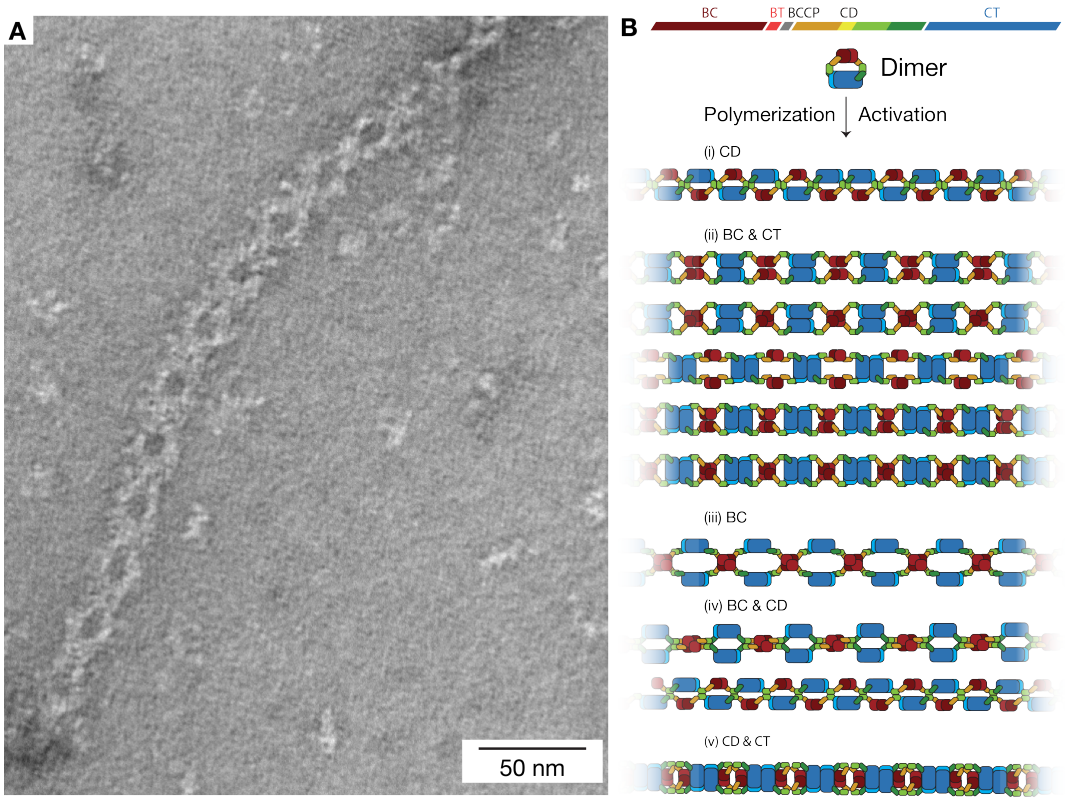


Figure 5.5 Conceptual models of filamentous human ACC.
A. Negative-stain electron micrograph of a human ACC1 polymer (unpublished data). **B.** Schematic of various possible filament architectures, sorted by their base of interaction (see main text). Coloring according to the sequence scheme. For simplicity, BT, BCCP and CD_L are omitted.

The mechanisms of other regulatory interactions of human ACC, including the binding of the regulatory protein MIG12[84], as well as allosteric inhibition by palmitate or allosteric activation by citrate, remain elusive[187]. MIG12 is involved in filament formation; possibly MIG12 binding induces a conformational change or shifts a conformational equilibrium and thus primes the ACC for polymerization.

So far, no allosteric citrate binding sites have been observed in structures of isolated BC or CT domains of eukaryotic ACCs. Citrate may indeed interact with the CD, and it is tempting to speculate that the CD_{C1}/CD_{C2} interface, which binds the regulatory loop in yeast ACC and is enriched in arginine residues in all eukaryotic ACCs, may also provide a binding site for citrate. Citrate binding to arginines would occupy the binding sites without inducing a stabilized interdomain interface and thus could plausibly lead to an activation of ACC.

5.4 *D. radiodurans* YCC is a dynamic multienzyme carboxylase

Besides the hybrid structure of *D. radiodurans* YCC presented above, only one additional structure, Map LCC, which has been published during the course of this work, is available for a bacterial multienzyme acyl-CoA carboxylase[86]. The two structures reflect a common, highly dynamic architecture, with distinct specializations. The crystal structure of Map LCC revealed a rigid CT core flanked laterally with dimeric BC domains, which are partially flexible based on EM analysis. The same type of conformational freedom, but with even larger amplitudes due to longer BC-BCCP linkers, is observed and partially quantified in our work on Dra YCC by SAXS and EM experiments.

Even though YCC shares the domain composition of BC, BCCP and CT with other bacterial acyl-CoA carboxylases, such as MCC and PCC, it exhibits a completely distinct architecture. Two main differences can be identified on a sequence level. First, YCC is lacking a BT domain, that mediates contacts between BC and CT in MCC and PCC, and second, BCCP is double-tethered at both termini due to the multienzyme nature of YCC. In other biotin-dependent carboxylases, except eukaryotic ACC, BCCP is exclusively located at the end of a polypeptide chain and therefore only single-tethered.

Although the active site cleft is partially occluded in Map LCC, this enzyme prefers long-chain acyl-CoA substrates although it is active towards acyl-CoA chains ranging from C_2 to C_{16}. Based on the active site architecture of the CT with its elongated substrate tunnel, the genomic synteny with an enzyme categorized as very-long-chain fatty acyl-CoA ligase, Dra YCC could act on long-chain or very-long chain acyl-CoAs. In preliminary activity assays, however, no activity could be detected for Dra YCC towards fully-saturated linear C_{16}/C_{18} substrates. Possibly, Dra YCC is active on distinct specialized long-chain acyl

substrates, which were not included in the initial assays. It might be less promiscuous than Map LCC with higher specificity for a unique substrate specific to *D. radiodurans* or related species. Additional activity assays including *D. radiodurans* whole cell extract, will be performed in the future to further characterize this multienzyme. Whether the increased flexibility is linked to overall activity or specificity is currently not known.

5.5 Dynamics of biotin-dependent carboxylase multienzymes

As discussed in the previous three sections, yeast ACC and Dra YCC are characterized by distinct architectures, but share an outstanding extent of conformational dynamics. The only two acyl-CoA carboxylases with available structures, besides yeast ACC, Dra YCC and Map LCC, are the multi-subunit enzymes PCC and MCC[58,59]. A comparable degree of conformational variability cannot be detected in neither of the two, as both are rigid enzyme complexes with limited large-scale flexibility, with the exception of the mobile BCCP. This observation is based not only on the analysis of static crystal structures, but also using negative stain EM, which reveals a high-level of conformational homogeneity for PCC and MCC. Thus, based on the available data for acyl-CoA carboxylases, the dynamic architecture is a unique feature of multienzymes.

More extensive structural data is available for biotin-dependent carboxylases acting on non-acyl-CoA substrates. In PC, a comparable conformational freedom has not been reported, although divergent overall architectures have been depicted. *Staphylococcus aureus* PC (SaPC) and human PC are symmetrical tetramers[61], formed by dimerization of dimers, while PC from *Rhizobium etli* (RePC) exhibits an asymmetric architecture[60], even though it is also built by a dimer of dimers (Figure 5.6). Initially, it was proposed that the distinct architectures might represent different states of the enzyme along the reaction coordinate, since RePC was crystallized in the presence of an activator, acetyl-CoA. This hypothesis was falsified later by x-ray crystallographic and EM analyses revealing that SaPC retains its symmetric organization also when acetyl-CoA is bound[346], and that the overall shape of RePC is asymmetric also without bound activator[347]. Thus the architectural difference between these SaPC and RePC is not representing different states of one enzyme, but is rather a further

example for varying architectures encountered frequently in biotin-dependent carboxylases, even within families of closely related enzymes.

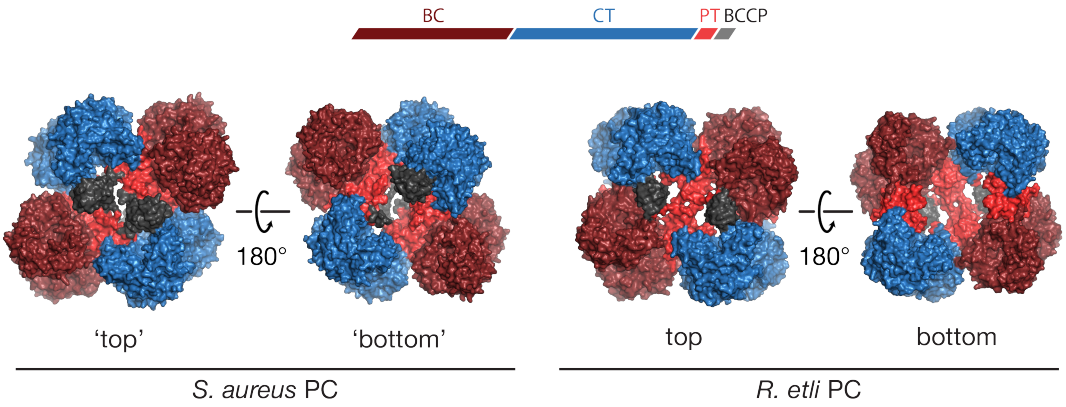


Figure 5.6 Different overall architectures for prokaryotic PC.
Surface representations of PC from *Staphylococcus aureus* (SaPC, pdb: 2QF7) and *Rhizobium etli* PC (RePC, pdb: 3G8C). SaPC is a symmetric tetramer, except for the mobile BCCP, with the same architecture of BC, BT and CT in the top and the bottom layer (one BCCP in 'bottom' layer not resolved in the electron density). RePC shows asymmetry between the top and the bottom layer, exemplified in the positioning of the PT domain. BCCP domains of RePC are only defined in the top layer. All coloring according to the sequence scheme.

The UC component of urea amidolyase from *K. lactis* is, in contrast to all other biotin-dependent carboxylases, catalytically active as a monomer, and there is no evidence for large-scale movement of the domains relative to each other, except for the BCCP translocation[57].

In summary, divergent architectures are common in the family of prokaryotic multienzyme carboxylases, and further examples of this may still await. However, the dynamic architecture observed in yeast ACC and Dra YCC is unique amongst biotin-dependent carboxylases. It may well be linked to their unique multienzyme organization where all enzymatic subunits are integrated into a single type of polypeptide chain together with a double-tethered BCCP. In particular the carrier protein tethering may require additional conformational freedom to compensate for a more restricted range of BCCP mobility.

5.6 Implications for carrier protein-dependent enzymes

Extensive conformational flexibility is not a common feature of large or oligomeric enzyme systems. However, it has previously been observed in carrier protein-dependent enzyme systems, such as the FAS and PKS multienzymes as well as

the giant multi-subunit PDC, and is now also confirmed for BCCP-dependent multienzyme acyl-CoA carboxylases. Previously, substrate transfer in carrier protein based enzyme systems has commonly been discussed only in terms of the mobility of the carrier protein. However, the cumulative data on the organization of carrier protein-dependent multienzymes now suggest an important role of large-scale conformational dynamics beyond the mobile carrier protein in substrate transfer and catalysis.

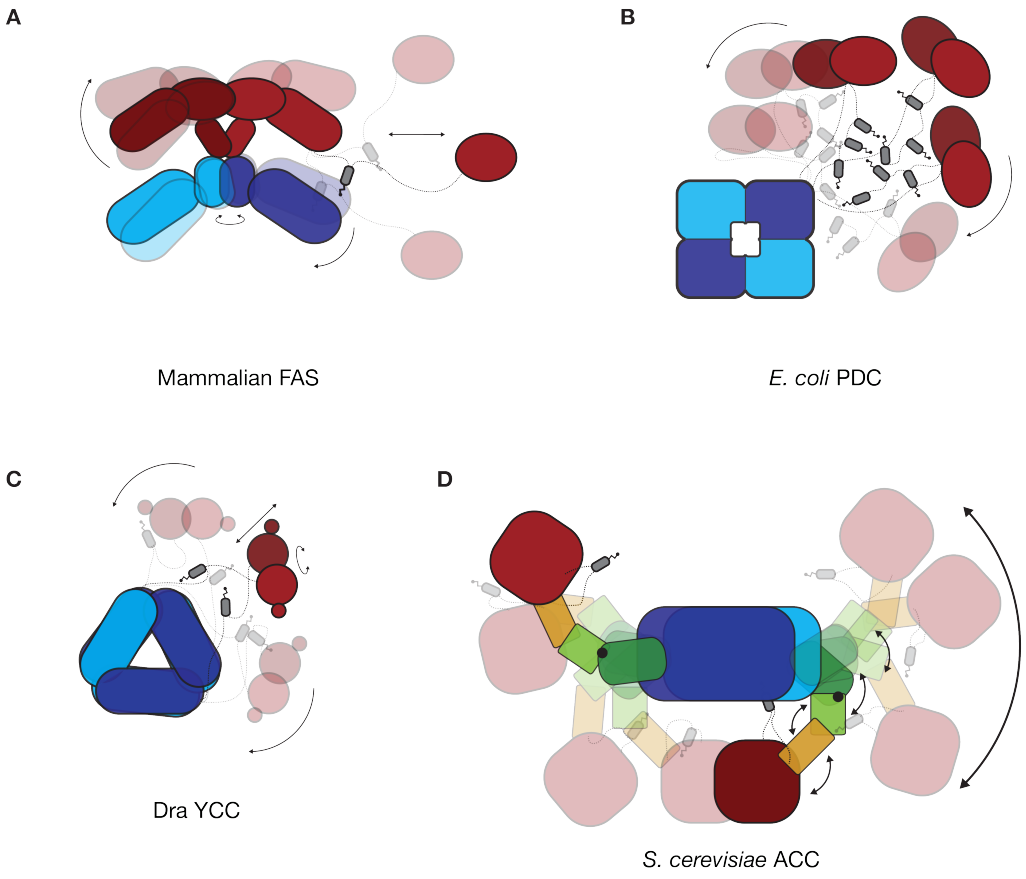


Figure 5.7 Dynamics of carrier protein-dependent multienzymes.
This figure is adapted and expanded from Figure 3.6. Transparent domains show the approximate range of motion of the domains. Arrows represent the possible domain movements. **A.** Schematic representation mammalian FAS with acyl-carrier protein, thioesterase domain and the linker tethering them the enzyme[19] **B.** Schematic representation of the *E. coli* pyruvate dehydrogenase complex. Only three E3 subunits are shown for clarity. Distances according to Murphy et al[20]. **C.** Schematic representation of Dra YCC containing only one set of BC and BCCP. **D.** Schematic representation of the dimeric *S. cerevisiae* ACC demonstrating the large-scale movements between the elongated, inactive architecture and the triangular, active conformation. BT and CD_L are not shown. The location where the regulatory loop can wedge between CD_{C1} and CD_{C2} is depicted with black dots in two instances of the CD.

Mammalian FAS and the related PKS undergo large rotational and hinge-like conformational changes during catalysis in order for ACP to reach all active sites,

and coupling between conformational and catalytic states has been proposed[26,348]. Interestingly, fungal FAS adopts a completely different and rather rigid architecture, which relies on a rigid scaffolding matrix. This altered architecture goes hand-in-hand with a highly specialized and unique variant of the ACP carrier protein as well as a rigidified ACP linker structure[19,349,350]. Presumably, these specializations establish a highly optimized rotational path for ACP-based substrate delivery even in the rigid fungal FAS barrel structure[349]. The fact that two fundamentally different architectures, where one is highly flexible and the other one is rigid, emerged for the same multi-step reaction and using the same set of domains, is reminiscent of variations of architectures between multi-subunit and multienzyme carboxylases, which also differ in ACP tethering.

The gigantic multi-subunit PDC employs a highly flexible and mobile architecture, with a rigid E2 core and flexibly tethered peripheral domains (E1, E3), with an allowed range of motion of approximately 150 Å. During catalysis, the lipoyl domain shuttles between these tethered domains[351].

Conceptually, the organization of Dra YCC resembles those of PDC, with one or more carrier protein domains shuttling between a rigid, symmetric core and highly mobile peripheral domains, which may reside at a distance to the core of up to 130 Å. In both, PDC and YCC, the large domain motions complement the mobility of the carrier protein. In PDC, however, the lipoyl domain carrier protein is only tethered on one side, N-terminally to the core-forming protein, and the flexible connection between the core and the peripheral units is provided by non-covalent interactions of the linker with binding domains.

The architecture of yeast ACC revealed in our and concomitant work however is truly unique: Yeast ACC employs a highly structured and functional 80 kDa CD linker region for connecting the BC-BCCP didomain to the dimeric CT core. Amongst the families of carrier protein-dependent multienzymes, the CD is by far the most complex linking region and comprises at least three distinct domains. Resolving multiple conformations of yeast ACC (fragments) at high-resolution additionally uncovers variable interactions within the CD and between the CD and catalytic domains in detail. These data indicate, that the CD is not used for a simple scaffolding-type positioning of catalytic domains, but has an active role in providing a series of flexible hinges governing large-scale movements that are linked to catalytic activity.

A comparison of high-resolution crystal structures of non-phosphorylated and authentically phosphorylated CD, demonstrates a unique regulatory mechanism of yeast ACC based on the restriction of hinge-bending motions in the CD. To our best knowledge, such a “mechanical” phosphorylation-based control mechanism has not been described for any other (carrier protein-based) multienzyme. However, the detailed role of confirmed phosphorylation sites in linker regions e.g. observed in animal FAS has not been uncovered so far. The current results on yeast ACC may thus well guide and inspire the search for novel linker-based regulatory mechanisms in animal FAS and other multienzymes.

5.7 Prospects for synthetic biology

Multienzymes are an attractive target for synthetic biology and combinatorial biosynthesis due to their numerous advantages over single enzymes, such as the prevention of side reactions, the stoichiometric presence and the common regulation of expression and activity. Researchers strive to engineer multienzymes in order to produce biologically active and relevant compounds. By swapping domains on a sequence level the outcome of the reaction catalyzed by the respective multienzyme can be altered. This was already successful, for example in the case of the PKS deoxyerythronolide B synthase, that was engineered to produce over 50 new unnatural compounds[40]. However, it is a difficult task since domain boundaries, domain interfaces and the dynamics of a system cannot be predicted with absolute certainty.

Interdomain connectivity is yet another feature that can often not directly be derived from the sequence, as seen for MCC, PCC and Map LCC, where extensive domain swaps happen in the core CT region, possibly depending on the nature of the substrate[58,59,86]. In addition, in many biotin-dependent carboxylases, the carrier protein from one protomer serves the active sites of another protomer, explaining why these enzymes are only active as oligomers. For such an intricate and flexible network to work properly, it is crucial that all linkers are of the appropriate length and that domains are positioned in exactly the right way relative to each other. In some cases, the ‘simple’ swapping of whole domains may thus not be sufficient without considering domain interactions, inter-domain connectivity and orientation.

In light of the increasing number of atomic structures of full-length multienzymes it becomes evident that dynamic architecture and large-range conformational changes are an integral functional aspect of many of these enzyme complexes. In order to modify biosynthetic pathways, there is a requirement to not only engineer the domain arrangement, but to also engineer the correct dynamics, in order to ensure full functionality of a multienzyme complex. Such requirement for overall conformational changes in multienzymes with a large number of distinct functional domains, such as animal FAS and PKS, is obvious and has been documented. The current work now reveals that even in systems combining only two distinct catalytic activities, large-scale conformational dynamics are an inherent part of the catalytic process. Thus, even for the design of simple two-step enzymatic systems based on carrier protein mediated substrate transfer, proper engineering of linking regions may be critical. Further detailed studies will be required to provide a more comprehensive description of the dynamics of linking regions, their interactions and conformational equilibria also in other multienzyme system.

The architecture of yeast ACC CD provides an exciting paradigm for engineering linker-regions with dual functionality. Conformational switching based on post-translational or chemical modification may be employed to control engineered enzymatic systems. Such linker-based mechanical control mechanisms enable a series of combinatorial control modes already in simple enzymatic systems, including (i) turning a two-enzyme system on or off (ii) uncoupling a two-enzyme system (iii) converting a three enzyme in a two-enzyme system (iv) switching between two synthetic routes in three-enzyme systems. Such control options open new prospects for combinatorial biosynthesis, which so far mostly relied on direct modification or swapping of catalytic domains, and may yield more versatile assemblies. Such novel engineered systems may ultimately contribute to unlocking the scientific and economical potential of synthetic biology and combinatorial biosynthesis[37].

6 Acknowledgements

First, I would like to thank you, Timm, for giving me the amazing opportunity and the trust to work in your lab on such a relevant and fascinating topic. Thank you for your permanent support, effort and time you invested in me. Thank you for being a friend. Your enthusiasm still amazes and inspires me.

Thanks to you, Sebastian and Markus, for being a part of my PhD advisory committee, I enjoyed the numerous committee meetings we had. Once a year, exactly according to protocol.

Special thanks to my oldest scientific companion, Imsi, for being a great friend over so many years. I hope you remember our PostDoc deal. And our squash deal. And our LaserTag deal. And our Las Vegas deal. And our Australia deal. I am sure we can increase this list in the next several months.

Thanks Cedi, my brother from another mother, of course you knew already that I would mention you here, didn't you?

Thanks to fabulous Fab, always ready for a beer. I will miss all the time we spent at the Türke or somewhere else during our nights out.

Many thanks go to the other members of the ACC team, crazy Eddi and Anna. It was always nice chatting about science, or football, or tennis, or curling, or basketball, or ice hockey, or biathlon, or swimming, or Olympic weight lifting, or the bachelor, or snowboarding, or skiing. I really appreciate all the time we spent together, whether it was in the lab, on the slopes, or in Bautzen.

Also special thanks to Roman, we know each other for a long time now, but still you have not become tired of my questions (I think so, at least..). Always helpful, always funny, always constructive, that's our RoRo.

Big thanks to all the rest of the Maier Lab, past and present, for making these five years go by so fast.

Thanks to the university of Basel, especially the Biozentrum, for supporting me.

Thanks to my family and special greetings to Emil (in many years I will show you this!).

7 References

1. Voet, D. & Voet, J.G. *Biochemistry*, xxv, 1428, 53 p. (John Wiley & Sons, Hoboken, NJ, 2011).
2. Schomburg, I. et al. BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic Acids Res* **41**, D764-72 (2013).
3. Srere, P.A. Complexes of sequential metabolic enzymes. *Annu Rev Biochem* **56**, 89-124 (1987).
4. Perham, R.N. SWINGING ARMS AND SWINGING DOMAINS IN MULTIFUNCTIONAL ENZYMES: Catalytic Machines for Multistep Reactions. *Annual Reviews of Biochemistry* **69**, 961-1004 (2000).
5. Graham, L.D., Packman, L.C. & Perham, R.N. Kinetics and specificity of reductive acylation of lipoyl domains from 2-oxo acid dehydrogenase multienzyme complexes. *Biochemistry* **28**, 1574-81 (1989).
6. Perham, R.N. Self-Assembly of Biological Macromolecules. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **272**, 123-136 (1975).
7. Roujeinikova, A. et al. Structural studies of fatty acyl-(acyl carrier protein) thioesters reveal a hydrophobic binding cavity that can expand to fit longer substrates. *J Mol Biol* **365**, 135-45 (2007).
8. Perham, R.N. Domains, motifs, and linkers in 2-oxo acid dehydrogenase multienzyme complexes: a paradigm in the design of a multifunctional protein. *Biochemistry* **30**, 8501-8512 (1991).
9. Reed, L.J. & Hackert, M.L. Structure-function relationships in dihydrolipoamide acyltransferases. *J Biol Chem* **265**, 8971-4 (1990).
10. Knowles, J.R. The Mechanism of Biotin-Dependent Enzymes. *Annual Reviews of Biochemistry* **58**, 195-221 (1989).
11. Hawkins, A.R. & Lamb, H.K. The Molecular Biology of Multidomain Proteins Selected Examples. *European Journal of Biochemistry* **232**, 7-18 (1995).
12. Nooren, I.M. & Thornton, J.M. Diversity of protein-protein interactions. *EMBO J* **22**, 3486-92 (2003).
13. Ekman, D., Björklund, Å.K., Frey-Skött, J. & Elofsson, A. Multi-domain Proteins in the Three Kingdoms of Life: Orphan Domains and Other Unassigned Regions. *Journal of Molecular Biology* **348**, 231-243 (2005).
14. Simmer, J.P., Kelly, R.E., Rinker, A.G., Jr., Scully, J.L. & Evans, D.R. Mammalian carbamyl phosphate synthetase (CPS). DNA sequence and evolution of the CPS domain of the Syrian hamster multifunctional protein CAD. *J Biol Chem* **265**, 10395-402 (1990).
15. Sumper, M., Oesterhelt, D., Riepertinger, C. & Lynen, F. [Synthesis of various carboxylic acids by the fatty acid synthetase multienzyme complex of yeast and the explanation for their structure]. *Eur J Biochem* **10**, 377-87 (1969).
16. Green, J.D., Perham, R.N., Ullrich, S.J. & Appella, E. Conformational studies of the interdomain linker peptides in the dihydrolipoyl acetyltransferase component of the pyruvate dehydrogenase multienzyme complex of Escherichia coli. *J Biol Chem* **267**, 23484-8 (1992).

17. Radford, S.E., Laue, E.D., Perham, R.N., Martin, S.R. & Appella, E. Conformational flexibility and folding of synthetic peptides representing an interdomain segment of polypeptide chain in the pyruvate dehydrogenase multienzyme complex of Escherichia coli. *J Biol Chem* **264**, 767-75 (1989).
18. Jenni, S., Leibundgut, M., Maier, T. & Ban, N. Architecture of a Fungal Fatty Acid Synthase at 5 A Resolution. *Science* **311**, 1263-1267 (2006).
19. Maier, T., Leibundgut, M., Boehringer, D. & Ban, N. Structure and function of eukaryotic fatty acid synthases. *Quarterly Reviews of Biophysics* **43**, 373-422 (2010).
20. Murphy, G.E. & Jensen, G.J. Electron Cryotomography of the E. coli Pyruvate and 2-Oxoglutarate Dehydrogenase Complexes. *Structure* **13**, 1765-1773 (2005).
21. de Kok, A., Hengeveld, A.F., Martin, A. & Westphal, A.H. The pyruvate dehydrogenase multienzyme complex from Gram-negative bacteria. *Biochim Biophys Acta* **1385**, 353-66 (1998).
22. Izard, T. et al. Principles of quasi-equivalence and Euclidean geometry govern the assembly of cubic and dodecahedral cores of pyruvate dehydrogenase complexes. *Proc Natl Acad Sci U S A* **96**, 1240-5 (1999).
23. Elovson, J. & Vagelos, P.R. Acyl carrier protein. X. Acyl carrier protein synthetase. *J Biol Chem* **243**, 3603-11 (1968).
24. Weber, T., Baumgartner, R., Renner, C., Marahiel, M.A. & Holak, T.A. Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases. *Structure* **8**, 407-418 (2000).
25. Maier, T., Leibundgut, M. & Ban, N. The crystal structure of a mammalian fatty acid synthase. *Science* **321**, 1315-22 (2008).
26. Brignole, E.J., Smith, S. & Asturias, F.J. Conformational flexibility of metazoan fatty acid synthase enables catalysis. *Nat Struct Mol Biol* **16**, 190-7 (2009).
27. Asturias, F.J. et al. Structure and molecular organization of mammalian fatty acid synthase. *Nat Struct Mol Biol* **12**, 225-32 (2005).
28. Ploskon, E. et al. A mammalian type I fatty acid synthase acyl carrier protein domain does not sequester acyl chains. *J Biol Chem* **283**, 518-28 (2008).
29. Ahlert, J. The Calicheamicin Gene Cluster and Its Iterative Type I Enediyne PKS. *Science* **297**, 1173-1176 (2002).
30. Aparicio, J.F. et al. Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hygroscopicus: Analysis of the enzymatic domains in the modular polyketide synthase. *Gene* **169**, 9-16 (1995).
31. Solenberg, P.J. et al. Production of hybrid glycopeptide antibiotics in vitro and in Streptomyces toyocaensis. *Chemistry & Biology* **4**, 195-202 (1997).
32. Walsh, C.T. The chemical versatility of natural-product assembly lines. *Acc Chem Res* **41**, 4-10 (2008).
33. Turgay, K., Krause, M. & Marahiel, M.A. Four homologous domains in the primary structure of GrsB are related to domains in a superfamily of adenylate-forming enzymes. *Mol Microbiol* **6**, 529-46 (1992).
34. Tong, L. Structure and function of biotin-dependent carboxylases. *Cellular and Molecular Life Sciences* **70**, 863-891 (2012).
35. Charusanti, P. et al. Exploiting adaptive laboratory evolution of Streptomyces clavuligerus for antibiotic discovery and overproduction. *PLoS One* **7**, e33727 (2012).

36. Peltier, K. The new macrolactones: clarithromycin (Biaxin) and azithromycin (Zithromax). *Conn Med* **56**, 371-4 (1992).
37. Weissman, K.J. Genetic engineering of modular PKSs: from combinatorial biosynthesis to synthetic biology. *Nat Prod Rep* (2015).
38. Giessen, T.W. & Marahiel, M.A. Ribosome-independent biosynthesis of biologically active peptides: Application of synthetic biology to generate structural diversity. *FEBS Letters* **586**, 2065-2075 (2012).
39. Hans, M., Hornung, A., Dziarnowski, A., Cane, D.E. & Khosla, C. Mechanistic Analysis of Acyl Transferase Domain Exchange in Polyketide Synthase Modules. *Journal of the American Chemical Society* **125**, 5366-5374 (2003).
40. McDaniel, R. et al. Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. *Proceedings of the National Academy of Sciences* **96**, 1846-1851 (1999).
41. Sundermann, U. et al. Enzyme-Directed Mutasynthesis: A Combined Experimental and Theoretical Approach to Substrate Recognition of a Polyketide Synthase. *ACS Chemical Biology* **8**, 443-450 (2013).
42. Cronan, J., John E & Waldrop, G.L. Multi-subunit acetyl-CoA carboxylases. *Progress in Lipid Research* **41**, 407-435 (2002).
43. Jitrapakdee, S. et al. Structure, mechanism and regulation of pyruvate carboxylase. *Biochemical Journal* **413**, 369-387 (2008).
44. Tong, L. Acetyl-coenzyme A carboxylase: crucial metabolic enzyme and attractive target for drug discovery. *Cell Mol Life Sci* **62**, 1784-803 (2005).
45. Wakil, S.J., Stoops, J.K. & Joshi, V.C. Fatty acid synthesis and its regulation. *Annu Rev Biochem* **52**, 537-79 (1983).
46. Berg, I.A. et al. Autotrophic carbon fixation in archaea. *Nat Rev Microbiol* **8**, 447-60 (2010).
47. Pratscher, J., Dumont, M.G. & Conrad, R. Ammonia oxidation coupled to CO2 fixation by archaea and bacteria in an agricultural soil. *Proc Natl Acad Sci U S A* **108**, 4170-5 (2011).
48. Smejkalova, H., Erb, T.J. & Fuchs, G. Methanol assimilation in Methylobacterium extorquens AM1: demonstration of all enzymes and their regulation. *PLoS One* **5**(2010).
49. Alber, B.E. Biotechnological potential of the ethylmalonyl-CoA pathway. *Appl Microbiol Biotechnol* **89**, 17-25 (2011).
50. Erb, T.J. et al. Synthesis of C5-dicarboxylic acids from C2-units involving crotonyl-CoA carboxylase/reductase: the ethylmalonyl-CoA pathway. *Proc Natl Acad Sci U S A* **104**, 10631-6 (2007).
51. Khomyakova, M., Bukmez, O., Thomas, L.K., Erb, T.J. & Berg, I.A. A methylaspartate cycle in haloarchaea. *Science* **331**, 334-7 (2011).
52. Schneider, K., Asao, M., Carter, M.S. & Alber, B.E. Rhodobacter sphaeroides uses a reductive route via propionyl coenzyme A to assimilate 3-hydroxypropionate. *J Bacteriol* **194**, 225-32 (2012).
53. Zhang, H., Boghigian, B.A. & Pfeifer, B.A. Investigating the role of native propionyl-CoA and methylmalonyl-CoA metabolism on heterologous polyketide production in Escherichia coli. *Biotechnol Bioeng* **105**, 567-73 (2010).
54. Navarathna, D.H., Harris, S.D., Roberts, D.D. & Nickerson, K.W. Evolutionary aspects of urea utilization by fungi. *FEMS Yeast Res* **10**, 209-13 (2010).

55. Gago, G., Diacovich, L., Arabolaza, A., Tsai, S.C. & Gramajo, H. Fatty acid biosynthesis in actinomycetes. *FEMS Microbiol Rev* **35**, 475-97 (2011).
56. Forster-Fromme, K. & Jendrossek, D. Catabolism of citronellol and related acyclic terpenoids in pseudomonads. *Appl Microbiol Biotechnol* **87**, 859-69 (2010).
57. Fan, C., Chou, C.Y., Tong, L. & Xiang, S. Crystal Structure of Urea Carboxylase Provides Insights into the Carboxyltransfer Reaction. *The Journal of Biological Chemistry* **287**, 9389-9398 (2012).
58. Huang, C.S., Ge, P., Zhou, Z.H. & Tong, L. An unanticipated architecture of the 750-kDa alpha6beta6 holoenzyme of 3-methylcrotonyl-CoA carboxylase. *Nature* **481**, 219-23 (2012).
59. Huang, C.S. et al. Crystal structure of the alpha(6)beta(6) holoenzyme of propionyl-coenzyme A carboxylase. *Nature* **466**, 1001-5 (2010).
60. St Maurice, M. et al. Domain architecture of pyruvate carboxylase, a biotin-dependent multifunctional enzyme. *Science* **317**, 1076-9 (2007).
61. Xiang, S. & Tong, L. Crystal structures of human and Staphylococcus aureus pyruvate carboxylase and molecular insights into the carboxyltransfer reaction. *Nat Struct Mol Biol* **15**, 295-302 (2008).
62. Mobley, H.L. & Hausinger, R.P. Microbial ureases: significance, regulation, and molecular characterization. *Microbiological Reviews* **53**, 85-108 (1989).
63. Sirko, A. & Brodzik, R. Plant ureases: roles and regulation. *Acta Biochim Pol* **47**, 1189-95 (2000).
64. Strope, P.K., Nickerson, K.W., Harris, S.D. & Moriyama, E.N. Molecular evolution of urea amidolyase and urea carboxylase in fungi. *BMC Evol Biol* **11**, 80 (2011).
65. Ghosh, S. et al. Arginine-induced germ tube formation in Candida albicans is essential for escape from murine macrophage line RAW 264.7. *Infect Immun* **77**, 1596-605 (2009).
66. Utter, M.F. & Keech, D.B. Formation of oxaloacetate from pyruvate and carbon dioxide. *J Biol Chem* **235**, PC17-8 (1960).
67. Attwood, P.V. The structure and the mechanism of action of pyruvate carboxylase. *Int J Biochem Cell Biol* **27**, 231-49 (1995).
68. Jitrapakdee, S., Vidal-Puig, A. & Wallace, J.C. Anaplerotic roles of pyruvate carboxylase in mammalian tissues. *Cell Mol Life Sci* **63**, 843-54 (2006).
69. Jitrapakdee, S. & Wallace, J.C. Structure, function and regulation of pyruvate carboxylase. *Biochem J* **340 (Pt 1)**, 1-16 (1999).
70. Robinson, B.H. Lactic acidemia and mitochondrial disease. *Mol Genet Metab* **89**, 3-13 (2006).
71. Aguilar, J.A. et al. Substrate Specificity of the 3-Methylcrotonyl Coenzyme A (CoA) and Geranyl-CoA Carboxylases from Pseudomonas aeruginosa. *Journal of Bacteriology* **190**, 4888-4893 (2008).
72. Guan, X., Diez, T., Prasad, T.K., Nikolau, B.J. & Wurtele, E.S. Geranoyl-CoA Carboxylase: A Novel Biotin-Containing Enzyme in Plants. *Archives of Biochemistry and Biophysics* **362**, 12-21 (1999).
73. Scholl-Burgi, S., Sass, J.O., Zschocke, J. & Karall, D. Amino acid metabolism in patients with propionic acidaemia. *J Inherit Metab Dis* **35**, 65-70 (2012).
74. Haberlandt, E. et al. Epilepsy in Patients with Propionic Acidemia. *Neuropediatrics* **40**, 120-125 (2009).

75. Rigo, F.K. et al. Propionic acid induces convulsions and protein carbonylation in rats. *Neurosci Lett* **408**, 151-4 (2006).
76. Nikolau, B.J., Ohlrogge, J.B. & Wurtele, E.S. Plant biotin-containing carboxylases. *Arch Biochem Biophys* **414**, 211-22 (2003).
77. Desviat, L.R. et al. Functional analysis of MCCA and MCCB mutations causing methylcrotonylglycinuria. *Mol Genet Metab* **80**, 315-20 (2003).
78. Stadler, S.C. et al. Newborn screening for 3-methylcrotonyl-CoA carboxylase deficiency: population heterogeneity of MCCA and MCCB mutations and impact on risk assessment. *Hum Mutat* **27**, 748-59 (2006).
79. Athappilly, F.K. & Hendrickson, W.A. Structure of the biotinyl domain of acetyl-coenzyme A carboxylase determined by MAD phasing. *Structure* **3**, 1407-19 (1995).
80. Shi, S., Chen, Y., Siewers, V. & Nielsen, J. Improving production of malonyl coenzyme A-derived metabolites by abolishing Snf1-dependent regulation of Acc1. *MBio* **5**, e01130-14 (2014).
81. Waldrop, G.L., Holden, H.M. & Maurice, M.S. The enzymes of biotin dependent CO 2metabolism: What structures reveal about their reaction mechanisms. *Protein Science* **21**, 1597-1619 (2012).
82. Hall, P.R. Transcarboxylase 12S crystal structure: hexamer assembly and substrate binding to a multienzyme core. *The EMBO Journal* **22**, 2334-2347 (2003).
83. Hügler, M., Krieger, R.S., Jahn, M. & Fuchs, G. Characterization of acetyl-CoA/propionyl-CoA carboxylase in Metallosphaera sedula. *European Journal of Biochemistry* **270**, 736-744 (2003).
84. Kim, C.W. et al. Induced polymerization of mammalian acetyl-CoA carboxylase by MIG12 provides a tertiary level of regulation of fatty acid synthesis. *Proc Natl Acad Sci U S A* **107**, 9626-31 (2010).
85. Lee, C.-K. et al. Biotinoyl domain of human acetyl-CoA carboxylase: Structural insights into the carboxyl transfer mechanism. *Proteins: Structure, Function, and Bioinformatics* **72**, 613-624 (2008).
86. Tran, T.H. et al. Structure and function of a single-chain, multi-domain long-chain acyl-CoA carboxylase. *Nature* **518**, 120-4 (2015).
87. Shen, Y., Volrath, S.L., Weatherly, S.C., Elich, T.D. & Tong, L. A mechanism for the potent inhibition of eukaryotic acetyl-coenzyme A carboxylase by soraphen A, a macrocyclic polyketide natural product. *Mol Cell* **16**, 881-91 (2004).
88. Zhang, H., Yang, Z., Shen, Y. & Tong, L. Crystal structure of the carboxyltransferase domain of acetyl-coenzyme A carboxylase. *Science* **299**, 2064-7 (2003).
89. Wakil, S.J., Titchener, E.B. & Gibson, D.M. Evidence for the participation of biotin in the enzymic synthesis of fatty acids. *Biochim Biophys Acta* **29**, 225-6 (1958).
90. Harwood, H.J., Jr. Acetyl-CoA carboxylase inhibition for the treatment of metabolic syndrome. *Curr Opin Investig Drugs* **5**, 283-9 (2004).
91. Harwood, H.J., Jr. Treating the metabolic syndrome: acetyl-CoA carboxylase inhibition. *Expert Opin Ther Targets* **9**, 267-81 (2005).
92. Kim, K.H. Regulation of mammalian acetyl-coenzyme A carboxylase. *Annu Rev Nutr* **17**, 77-99 (1997).
93. Munday, M.R. Regulation of mammalian acetyl-CoA carboxylase. *Biochem Soc Trans* **30**, 1059-64 (2002).
94. Foster, D.W. The role of the carnitine system in human metabolism. *Ann N Y Acad Sci* **1033**, 1-16 (2004).

95. McGarry, J.D. & Brown, N.F. The mitochondrial carnitine palmitoyltransferase system. From concept to molecular analysis. *Eur J Biochem* **244**, 1-14 (1997).
96. Ramsay, R.R., Gandour, R.D. & van der Leij, F.R. Molecular enzymology of carnitine transfer and transport. *Biochim Biophys Acta* **1546**, 21-43 (2001).
97. Abu-Elheiga, L., Matzuk, M.M., Abo-Hashema, K.A. & Wakil, S.J. Continuous fatty acid oxidation and reduced fat storage in mice lacking acetyl-CoA carboxylase 2. *Science* **291**, 2613-6 (2001).
98. Abu-Elheiga, L., Oh, W., Kordari, P. & Wakil, S.J. Acetyl-CoA carboxylase 2 mutant mice are protected against obesity and diabetes induced by high-fat/high-carbohydrate diets. *Proc Natl Acad Sci U S A* **100**, 10207-12 (2003).
99. Diaz, F.J. et al. Acetyl-coenzyme A carboxylase alpha gene variations may be associated with the direct effects of some antipsychotics on triglyceride levels. *Schizophr Res* **115**, 136-40 (2009).
100. Abramson, H.N. The Lipogenesis Pathway as a Cancer Target. *Journal of Medicinal Chemistry* **54**, 5615-5638 (2011).
101. Beckers, A. et al. Chemical Inhibition of Acetyl-CoA Carboxylase Induces Growth Arrest and Cytotoxicity Selectively in Cancer Cells. *Cancer Research* **67**, 8180-8187 (2007).
102. Chajes, V. Acetyl-CoA Carboxylase α Is Essential to Breast Cancer Cell Survival. *Cancer Research* **66**, 5287-5294 (2006).
103. Olsen, A.M. et al. Fatty acid synthesis is a therapeutic target in human liposarcoma. *International Journal of Oncology* **36**, 1309-1314 (2010).
104. Wang, C., Rajput, S., Watabe, K., Liao, D.-F. & Cao, D. Acetyl-CoA carboxylase-a as a novel target for cancer therapy. *Frontiers in bioscience* **S2**, 515-26 (2010).
105. Al-Feel, W., DeMar, J.C. & Wakil, S.J. A Saccharomyces cerevisiae mutant strain defective in acetyl-CoA carboxylase arrests at the G2/M phase of the cell cycle. *Proc Natl Acad Sci U S A* **100**, 3095-100 (2003).
106. Sasaki, Y. & Nagano, Y. Plant acetyl-CoA carboxylase: structure, biosynthesis, regulation, and gene manipulation for plant breeding. *Biosci Biotechnol Biochem* **68**, 1175-84 (2004).
107. Gao, H., Sumanaweera, N., Bailer, S.M. & Stochaj, U. Nuclear accumulation of the small GTPase Gsp1p depends on nucleoporins Nup133p, Rat2p/Nup120p, Nup85p, Nic96p, and the acetyl-CoA carboxylase Acc1p. *J Biol Chem* **278**, 25331-40 (2003).
108. Hiltunen, J.K., Chen, Z., Haapalainen, A.M., Wierenga, R.K. & Kastaniotis, A.J. Mitochondrial fatty acid synthesis--an adopted set of enzymes making a pathway of major importance for the cellular metabolism. *Prog Lipid Res* **49**, 27-45 (2010).
109. Hoja, U. et al. HFA1 encoding an organelle-specific acetyl-CoA carboxylase controls mitochondrial fatty acid synthesis in Saccharomyces cerevisiae. *J Biol Chem* **279**, 21779-86 (2004).
110. Tehlivets, O., Scheuringer, K. & Kohlwein, S.D. Fatty acid synthesis and elongation in yeast. *Biochim Biophys Acta* **1771**, 255-70 (2007).
111. Diacovich, L. Kinetic and Structural Analysis of a New Group of Acyl-CoA Carboxylases Found in Streptomyces coelicolor A3(2). *The Journal of Biological Chemistry* **277**, 31228-31236 (2002).

112. Eberle, D., Hegarty, B., Bossard, P., Ferre, P. & Foufelle, F. SREBP transcription factors: master regulators of lipid homeostasis. *Biochimie* **86**, 839-48 (2004).
113. Field, F.J., Born, E., Murthy, S. & Mathur, S.N. Polyunsaturated fatty acids decrease the expression of sterol regulatory element-binding protein-1 in CaCo-2 cells: effect on fatty acid synthesis and triacylglycerol transport. *Biochem J* **368**, 855-64 (2002).
114. Foufelle, F. & Ferre, P. New perspectives in the regulation of hepatic glycolytic and lipogenic genes by insulin and glucose: a role for the transcription factor sterol regulatory element binding protein-1c. *Biochem J* **366**, 377-91 (2002).
115. Lin, J. et al. Hyperlipidemic effects of dietary saturated fats mediated through PGC-1beta coactivation of SREBP. *Cell* **120**, 261-73 (2005).
116. Puigserver, P. & Spiegelman, B.M. Peroxisome proliferator-activated receptor-gamma coactivator 1 alpha (PGC-1 alpha): transcriptional coactivator and metabolic regulator. *Endocr Rev* **24**, 78-90 (2003).
117. Barber, M.C., Price, N.T. & Travers, M.T. Structure and regulation of acetyl-CoA carboxylase genes of metazoa. *Biochim Biophys Acta* **1733**, 1-28 (2005).
118. Mao, J., Chirala, S.S. & Wakil, S.J. Human acetyl-CoA carboxylase 1 gene: presence of three promoters and heterogeneity at the 5'-untranslated mRNA region. *Proc Natl Acad Sci U S A* **100**, 7515-20 (2003).
119. Mao, J. & Seyfert, H.M. Promoter II of the bovine acetyl-coenzyme A carboxylase-alpha-encoding gene is widely expressed and strongly active in different cells. *Biochim Biophys Acta* **1576**, 324-9 (2002).
120. Oh, S.Y. et al. Alternative usages of multiple promoters of the acetyl-CoA carboxylase beta gene are related to differential transcriptional regulation in human and rodent tissues. *J Biol Chem* **280**, 5909-16 (2005).
121. Hasslacher, M., Ivessa, A.S., Paltauf, F. & Kohlwein, S.D. Acetyl-CoA carboxylase from yeast is an essential enzyme and is regulated by factors that control phospholipid metabolism. *J Biol Chem* **268**, 10946-52 (1993).
122. Shirra, M.K. et al. Inhibition of acetyl coenzyme A carboxylase activity restores expression of the INO1 gene in a snf1 mutant strain of Saccharomyces cerevisiae. *Mol Cell Biol* **21**, 5710-22 (2001).
123. Chirala, S.S., Zhong, Q., Huang, W. & al-Feel, W. Analysis of FAS3/ACC regulatory region of Saccharomyces cerevisiae: identification of a functional UASINO and sequences responsible for fatty acid mediated repression. *Nucleic Acids Res* **22**, 412-8 (1994).
124. Lenhard, J.M. & Gottschalk, W.K. Preclinical developments in type 2 diabetes. *Adv Drug Deliv Rev* **54**, 1199-212 (2002).
125. Berg, J., Tymoczenko, J.L. & Stryer, L. *Biochemistry*, (W. H. Freeman, 2002).
126. Davies, S.P., Sim, A.T. & Hardie, D.G. Location and function of three sites phosphorylated on rat acetyl-CoA carboxylase by the AMP-activated protein kinase. *Eur J Biochem* **187**, 183-90 (1990).
127. Ha, J., Daniel, S., Broyles, S.S. & Kim, K.H. Critical phosphorylation sites for acetyl-CoA carboxylase activity. *J Biol Chem* **269**, 22162-8 (1994).
128. Cho, Y.S. et al. Molecular mechanism for the regulation of human ACC2 through phosphorylation by AMPK. *Biochem Biophys Res Commun* **391**, 187-92 (2010).

129. Shen, Y. & Tong, L. Structural evidence for direct interactions between the BRCT domains of human BRCA1 and a phospho-peptide from human ACC1. *Biochemistry* **47**, 5767-73 (2008).
130. Williams, R.S., Lee, M.S., Hau, D.D. & Glover, J.N. Structural basis of phosphopeptide recognition by the BRCT domain of BRCA1. *Nat Struct Mol Biol* **11**, 519-25 (2004).
131. Ficarro, S.B. et al. Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nat Biotechnol* **20**, 301-5 (2002).
132. Woods, A. et al. Yeast SNF1 is functionally related to mammalian AMP-activated protein kinase and regulates acetyl-CoA carboxylase in vivo. *J Biol Chem* **269**, 19509-15 (1994).
133. Weatherly, S.C., Volrath, S.L. & Elich, T.D. Expression and characterization of recombinant fungal acetyl-CoA carboxylase and isolation of a soraphen-binding domain. *Biochem J* **380**, 105-10 (2004).
134. Shen, Y., Chou, C.Y., Chang, G.G. & Tong, L. Is dimerization required for the catalytic activity of bacterial biotin carboxylase? *Mol Cell* **22**, 807-18 (2006).
135. Diacovich, L. et al. Crystal structure of the beta-subunit of acyl-CoA carboxylase: structure-based engineering of substrate specificity. *Biochemistry* **43**, 14027-36 (2004).
136. Reaven, G.M. Insulin resistance: the link between obesity and cardiovascular disease. *Med Clin North Am* **95**, 875-92 (2011).
137. Wang, Y.C., McPherson, K., Marsh, T., Gortmaker, S.L. & Brown, M. Health and economic burden of the projected obesity trends in the USA and the UK. *Lancet* **378**, 815-25 (2011).
138. Abu-Elheiga, L. et al. Mutant mice lacking acetyl-CoA carboxylase 1 are embryonically lethal. *Proc Natl Acad Sci U S A* **102**, 12011-6 (2005).
139. Mao, J. et al. Liver-specific deletion of acetyl-CoA carboxylase 1 reduces hepatic triglyceride accumulation without affecting glucose homeostasis. *Proc Natl Acad Sci U S A* **103**, 8552-7 (2006).
140. Alkhateeb, H., Holloway, G.P. & Bonen, A. Skeletal muscle fatty acid oxidation is not directly associated with AMPK or ACC2 phosphorylation. *Appl Physiol Nutr Metab* **36**, 361-7 (2011).
141. Hoehn, K.L. et al. Acute or chronic upregulation of mitochondrial fatty acid oxidation has no net effect on whole-body energy expenditure or adiposity. *Cell Metab* **11**, 70-6 (2010).
142. Olson, D.P., Pulinilkunnil, T., Cline, G.W., Shulman, G.I. & Lowell, B.B. Gene knockout of Acc2 has little effect on body weight, fat mass, or food intake. *Proc Natl Acad Sci U S A* **107**, 7598-603 (2010).
143. Bengtsson, C. et al. Design of small molecule inhibitors of acetyl-CoA carboxylase 1 and 2 showing reduction of hepatic malonyl-CoA levels in vivo in obese Zucker rats. *Bioorg Med Chem* **19**, 3039-53 (2011).
144. Corbett, J.W. Review of recent acetyl-CoA carboxylase inhibitor patents: mid-2007-2008. *Expert Opin Ther Pat* **19**, 943-56 (2009).
145. Corbett, J.W. et al. Discovery of small molecule isozyme non-specific inhibitors of mammalian acetyl-CoA carboxylase 1 and 2. *Bioorg Med Chem Lett* **20**, 2383-8 (2010).
146. Harwood, H.J., Jr. et al. Isozyme-nonselective N-substituted bipiperidylcarboxamide acetyl-CoA carboxylase inhibitors reduce tissue malonyl-CoA concentrations, inhibit fatty acid synthesis, and increase

fatty acid oxidation in cultured cells and in experimental animals. *J Biol Chem* **278**, 37099-111 (2003).
147. Jonas, M., LaMarr, W.A. & Ozbal, C. Mass spectrometry in high-throughput screening: a case study on acetyl-coenzyme a carboxylase using RapidFire--mass spectrometry (RF-MS). *Comb Chem High Throughput Screen* **12**, 752-9 (2009).
148. Marjanovic, J. et al. Recombinant yeast screen for new inhibitors of human acetyl-CoA carboxylase 2 identifies potential drugs to treat obesity. *Proc Natl Acad Sci U S A* **107**, 9093-8 (2010).
149. Glien, M. et al. Stimulation of fat oxidation, but no sustained reduction of hepatic lipids by prolonged pharmacological inhibition of acetyl CoA carboxylase. *Horm Metab Res* **43**, 601-6 (2011).
150. Ronnebaum, S.M. et al. Chronic suppression of acetyl-CoA carboxylase 1 in beta-cells impairs insulin secretion via inhibition of glucose rather than lipid metabolism. *J Biol Chem* **283**, 14248-56 (2008).
151. Beckers, A. et al. Chemical inhibition of acetyl-CoA carboxylase induces growth arrest and cytotoxicity selectively in cancer cells. *Cancer Res* **67**, 8180-7 (2007).
152. Scott, K.E. et al. Metabolic regulation of invadopodia and invasion by acetyl-CoA carboxylase 1 and de novo lipogenesis. *PLoS One* **7**, e29761 (2012).
153. Seng, T.W., Skillman, T.R., Yang, N. & Hammond, C. Cyclohexanedione herbicides are inhibitors of rat heart acetyl-CoA carboxylase. *Bioorg Med Chem Lett* **13**, 3237-42 (2003).
154. Jakob, R.P. et al. Dimeric Structure of the Bacterial Extracellular Foldase PrsA. *J Biol Chem* **290**, 3278-92 (2015).
155. Fanghanel, J. & Fischer, G. Insights into the catalytic mechanism of peptidyl prolyl cis/trans isomerases. *Front Biosci* **9**, 3453-78 (2004).
156. Gothel, S.F. & Marahiel, M.A. Peptidyl-prolyl cis-trans isomerases, a superfamily of ubiquitous folding catalysts. *Cell Mol Life Sci* **55**, 423-36 (1999).
157. Rahfeld, J.U. et al. Confirmation of the existence of a third family among peptidyl-prolyl cis/trans isomerases. Amino acid sequence and recombinant production of parvulin. *FEBS Lett* **352**, 180-4 (1994).
158. Schmid, F.X. Prolyl isomerases. *Adv Protein Chem* **59**, 243-82 (2001).
159. Ranganathan, R., Lu, K.P., Hunter, T. & Noel, J.P. Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent. *Cell* **89**, 875-886 (1997).
160. Chen, H.I. & Sudol, M. The WW domain of Yes-associated protein binds a proline-rich ligand that differs from the consensus established for Src homology 3-binding modules. *Proc Natl Acad Sci U S A* **92**, 7819-23 (1995).
161. Yaffe, M.B. et al. Sequence-specific and phosphorylation-dependent proline isomerization: a potential mitotic regulatory mechanism. *Science* **278**, 1957-60 (1997).
162. Shen, M., Stukenberg, P.T., Kirschner, M.W. & Lu, K.P. The essential mitotic peptidyl-prolyl isomerase Pin1 binds and regulates mitosis-specific phosphoproteins. *Genes Dev* **12**, 706-20 (1998).
163. Wu, S.C., Ye, R., Wu, X.C., Ng, S.C. & Wong, S.L. Enhanced secretory production of a single-chain antibody fragment from Bacillus subtilis by coproduction of molecular chaperones. *J Bacteriol* **180**, 2830-5 (1998).

164. Sekerina, E. et al. NMR solution structure of hPar14 reveals similarity to the peptidyl prolyl cis/trans isomerase domain of the mitotic regulator hPin1 but indicates a different functionality of the protein. *J Mol Biol* **301**, 1003-17 (2000).
165. Rudd, K.E. et al. A new family of peptidyl-prolyl isomerases. *Trends Biochem Sci* **20**, 12-4 (1995).
166. Behrens-Kneip, S. The role of SurA factor in outer membrane protein transport and virulence. *Int J Med Microbiol* **300**, 421-8 (2010).
167. Kale, A. et al. The virulence factor PEB4 (Cj0596) and the periplasmic protein Cj1289 are two structurally related SurA-like chaperones in the human pathogen Campylobacter jejuni. *J Biol Chem* **286**, 21254-65 (2011).
168. Jacobs, M., Andersen, J.B., Kontinen, V. & Sarvas, M. Bacillus subtilis PrsA is required in vivo as an extracytoplasmic chaperone for secretion of active enzymes synthesized either with or without pro-sequences. *Mol Microbiol* **8**, 957-66 (1993).
169. Vitikainen, M. et al. Structure-function analysis of PrsA reveals roles for the parvulin-like and flanking N- and C-terminal domains in protein folding and secretion in Bacillus subtilis. *J Biol Chem* **279**, 19302-14 (2004).
170. Vitikainen, M. et al. Quantitation of the capacity of the secretion apparatus and requirement for PrsA in growth and secretion of alpha-amylase in Bacillus subtilis. *J Bacteriol* **183**, 1881-90 (2001).
171. Drouault, S. et al. The peptidyl-prolyl isomerase motif is lacking in PmpA, the PrsA-like protein involved in the secretion machinery of Lactococcus lactis. *Appl Environ Microbiol* **68**, 3932-42 (2002).
172. Guo, L. et al. Phenotypic characterization of the foldase homologue PrsA in Streptococcus mutans. *Mol Oral Microbiol* **28**, 154-65 (2013).
173. Hyyrylainen, H.L. et al. Penicillin-binding protein folding is dependent on the PrsA peptidyl-prolyl cis-trans isomerase in Bacillus subtilis. *Mol Microbiol* **77**, 108-27 (2010).
174. Alonzo, F., 3rd, Xayarath, B., Whisstock, J.C. & Freitag, N.E. Functional analysis of the Listeria monocytogenes secretion chaperone PrsA2 and its multiple contributions to bacterial virulence. *Mol Microbiol* **80**, 1530-48 (2011).
175. Jousselin, A. et al. The posttranslocational chaperone lipoprotein PrsA is involved in both glycopeptide and oxacillin resistance in Staphylococcus aureus. *Antimicrob Agents Chemother* **56**, 3629-40 (2012).
176. Williams, R.C. et al. Production of Bacillus anthracis protective antigen is dependent on the extracellular chaperone, PrsA. *J Biol Chem* **278**, 18056-62 (2003).
177. Alonzo, F., 3rd & Freitag, N.E. Listeria monocytogenes PrsA2 is required for virulence factor secretion and bacterial viability within the host cell cytosol. *Infect Immun* **78**, 4944-57 (2010).
178. Milgraum, L.Z., Witters, L.A., Pasternack, G.R. & Kuhajda, F.P. Enzymes of the fatty acid synthesis pathway are highly expressed in in situ breast carcinoma. *Clin Cancer Res* **3**, 2115-20 (1997).
179. Swinnen, J.V., Brusselmans, K. & Verhoeven, G. Increased lipogenesis in cancer cells: new players, novel targets. *Curr Opin Clin Nutr Metab Care* **9**, 358-65 (2006).
180. Campbell, J.W. & Cronan, J.E., Jr. Bacterial fatty acid biosynthesis: targets for antibacterial drug discovery. *Annu Rev Microbiol* **55**, 305-32 (2001).

181. Wakil, S.J. & Abu-Elheiga, L.A. Fatty acid metabolism: target for metabolic syndrome. *J Lipid Res* **50 Suppl**, S138-43 (2009).
182. Alberts, A.W. & Vagelos, P.R. Acetyl CoA carboxylase. I. Requirement for two protein fractions. *Proc Natl Acad Sci U S A* **59**, 561-8 (1968).
183. Alberts, A.W., Nervi, A.M. & Vagelos, P.R. Acetyl CoA carboxylase, II. Deomonstration of biotin-protein and biotin carboxylase subunits. *Proc Natl Acad Sci U S A* **63**, 1319-26 (1969).
184. Cronan, J.E., Jr. & Waldrop, G.L. Multi-subunit acetyl-CoA carboxylases. *Prog Lipid Res* **41**, 407-35 (2002).
185. Bianchi, A. et al. Identification of an isozymic form of acetyl-CoA carboxylase. *J Biol Chem* **265**, 1502-9 (1990).
186. Abu-Elheiga, L. et al. The subcellular localization of acetyl-CoA carboxylase 2. *Proc Natl Acad Sci U S A* **97**, 1444-9 (2000).
187. Brownsey, R.W., Boone, A.N., Elliott, J.E., Kulpa, J.E. & Lee, W.M. Regulation of acetyl-CoA carboxylase. *Biochem Soc Trans* **34**, 223-7 (2006).
188. Waldrop, G.L., Rayment, I. & Holden, H.M. Three-dimensional structure of the biotin carboxylase subunit of acetyl-CoA carboxylase. *Biochemistry* **33**, 10249-56 (1994).
189. Bilder, P. et al. The structure of the carboxyltransferase component of acetyl-coA carboxylase reveals a zinc-binding motif unique to the bacterial enzyme. *Biochemistry* **45**, 1712-22 (2006).
190. Tong, L. Structure and function of biotin-dependent carboxylases. *Cell Mol Life Sci* **70**, 863-91 (2013).
191. Ray, H., Moreau, K., Dizin, E., Callebaut, I. & Venezia, N.D. ACCA phosphopeptide recognition by the BRCT repeats of BRCA1. *J Mol Biol* **359**, 973-82 (2006).
192. Holt, L.J. et al. Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. *Science* **325**, 1682-6 (2009).
193. Diacovich, L. et al. Kinetic and structural analysis of a new group of Acyl-CoA carboxylases found in Streptomyces coelicolor A3(2). *J Biol Chem* **277**, 31228-36 (2002).
194. Wei, J. & Tong, L. Crystal structure of the 500-kDa yeast acetyl-CoA carboxylase holoenzyme dimer. *Nature* **526**, 723-727 (2015).
195. Bernado, P. Effect of interdomain dynamics on the structure determination of modular proteins by small-angle scattering. *Eur Biophys J* **39**, 769-80 (2010).
196. Smith, J.L., Skiniotis, G. & Sherman, D.H. Architecture of the polyketide synthase module: surprises from electron cryo-microscopy. *Curr Opin Struct Biol* **31**, 9-19 (2015).
197. Strieker, M., Tanovic, A. & Marahiel, M.A. Nonribosomal peptide synthetases: structures and dynamics. *Curr Opin Struct Biol* **20**, 234-40 (2010).
198. Patel, M.S., Nemeria, N.S., Furey, W. & Jordan, F. The pyruvate dehydrogenase complexes: structure-based function and regulation. *J Biol Chem* **289**, 16615-23 (2014).
199. Fitzgerald, D.J. et al. Protein complex expression by using multigene baculoviral vectors. *Nat Methods* **3**, 1021-32 (2006).
200. Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-32 (2010).
201. Zhang, K.Y., Cowtan, K. & Main, P. Combining constraints for electron-density modification. *Methods Enzymol* **277**, 53-64 (1997).

202. Winn, M.D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-42 (2011).
203. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-21 (2010).
204. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).
205. Sheldrick, G.M. A short history of SHELX. *Acta Crystallogr A* **64**, 112-22 (2008).
206. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica Section D-Biological Crystallography* **60**, 2256-2268 (2004).
207. Pape, T. & Schneider, T.R. HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *J Appl Crystallogr* **37**, 843-844 (2004).
208. Olsen, J.V. et al. Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci Signal* **3**, ra3 (2010).
209. Bian, Y. et al. An enzyme assisted RP-RPLC approach for in-depth analysis of human liver phosphoproteome. *J Proteomics* **96**, 253-62 (2014).
210. Cantin, G.T. et al. Combining protein-based IMAC, peptide-based IMAC, and MudPIT for efficient phosphoproteomic analysis. *J Proteome Res* **7**, 1346-51 (2008).
211. McCoy, A.J. et al. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).
212. Biasini, M. et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* **42**, W252-8 (2014).
213. Pei, J. & Grishin, N.V. AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics* **17**, 700-12 (2001).
214. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539 (2011).
215. Petoukhov, M.V. et al. New developments in the program package for small-angle scattering data analysis. *J Appl Crystallogr* **45**, 342-350 (2012).
216. Skou, S., Gillilan, R.E. & Ando, N. Synchrotron-based small-angle X-ray scattering of proteins in solution. *Nature Protocols* **9**, 1727-1739 (2014).
217. Jacques, D.A. & Trewhella, J. Small-angle scattering for structural biology-Expanding the frontier while avoiding the pitfalls. *Protein Science* **19**, 642-657 (2010).
218. Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J. & Svergun, D.I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J Appl Crystallogr* **36**, 1277-1282 (2003).
219. Fischer, H., Neto, M.D., Napolitano, H.B., Polikarpov, I. & Craievich, A.F. Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *Journal of Applied Crystallography* **43**, 101-109 (2010).
220. Scheres, S.H. A Bayesian view on cryo-EM structure determination. *J Mol Biol* **415**, 406-18 (2012).

221. Marabini, R. et al. Xmipp: An Image Processing Package for Electron Microscopy. *J Struct Biol* **116**, 237-40 (1996).
222. Wakil, S.J., Stoops, J.K. & Joshi, V.C. Fatty Acid Synthesis and its Regulation. *Annual Reviews of Biochemistry* **52**, 537-579 (1983).
223. Tong, L. Acetyl-coenzyme A carboxylase: crucial metabolic enzyme and attractive target for drug discovery. *Cellular and Molecular Life Sciences* **62**, 1784-1803 (2005).
224. Zhang, H., Boghigian, B.A. & Pfeifer, B.A. Investigating the role of native propionyl-CoA and methylmalonyl-CoA metabolism on heterologous polyketide production in Escherichia coli. *Biotechnology and Bioengineering* **105**, 567-573 (2010).
225. Maurice, M.S. et al. Domain Architecture of Pyruvate Carboxylase, a Biotin-Dependent Multifunctional Enzyme. *Science* **317**, 1076-1079 (2007).
226. Fall, R.R. & Hector, M.L. Acyl-coenzyme A carboxylases. Homologous 3-methylcrotonyl-CoA and geranyl-CoA carboxylases from Pseudomonas citronellolis. *Biochemistry* **16**, 4000-4005 (1977).
227. Rodriguez, E., Banchio, C., Diacovich, L., Bibb, M.J. & Gramajo, H. Role of an Essential Acyl Coenzyme A Carboxylase in the Primary and Secondary Metabolism of Streptomyces coelicolor. *Applied and Environmental Microbiology* **67**, 4166-4176 (2001).
228. Huang, C.S. et al. Crystal structure of the α 6 β 6 holoenzyme of propionyl-coenzyme A carboxylase. *Nature* **466**, 1001-1005 (2010).
229. Huang, C.S., Ge, P., Zhou, Z.H. & Tong, L. An unanticipated architecture of the 750-kDa α 6 β 6 holoenzyme of 3-methylcrotonyl-CoA carboxylase. *Nature* **481**, 219-223 (2011).
230. Attwood, P.V. & Wallace, J.C. Chemical and Catalytic Mechanisms of Carboxyl Transfer Reactions in Biotin-Dependent Enzymes. *Accounts of Chemical Research* **35**, 113-120 (2002).
231. Kim, C.-W. et al. Induced polymerization of mammalian acetyl-CoA carboxylase by MIG12 provides a tertiary level of regulation of fatty acid synthesis. *Proceedings of the National Academy of Sciences* **107**, 9626-9631 (2010).
232. Hunkeler, M., Stuttfeld, E., Hagmann, A., Imseng, S. & Maier, T. The dynamic organization of fungal acetyl-CoA carboxylase. *Nat Comms* **7**, 1-11 (2016).
233. Tran, T.H. et al. Structure and function of a single-chain, multi-domain long-chain acyl-CoA carboxylase. *Nature* **518**, 120-123 (2015).
234. Arabolaza, A. et al. Crystal Structures and Mutational Analyses of Acyl-CoA Carboxylase β Subunit of Streptomyces coelicolor. *Biochemistry* **49**, 7367-7376 (2010).
235. Reddy, M.C.M. et al. Structure, Activity, and Inhibition of the Carboxyltransferase-Subunit of Acetyl Coenzyme A Carboxylase (AccD6) from Mycobacterium tuberculosis. *Antimicrobial Agents and Chemotherapy* **58**, 6122-6132 (2014).
236. Kondo, S. et al. Structure of the biotin carboxylase domain of pyruvate carboxylase from Bacillus thermodenitrificans. *Acta Crystallographica Section D* **63**, 885-890 (2007).
237. Vagin, A.A. et al. REFMAC5 dictionary: organizatuon of prior chemical knowledge and guidelines for its use. *Acta Crystallographica Section D* **60**, 2184-2195 (2004).

238. Chou, C.Y., Yu, L.P.C. & Tong, L. Crystal Structure of Biotin Carboxylase in Complex with Substrates and Implications for Its Catalytic Mechanism. *The Journal of Biological Chemistry* **284**, 11690-11697 (2008).
239. Waldrop, G.L., Holden, H.M. & Maurice, M.S. The enzymes of biotin dependent CO2 metabolism: What structures reveal about their reaction mechanisms. *Protein Science* **21**, 1597-1619 (2012).
240. Thoden, J.B., Blanchard, C.Z., Holden, H.M. & Waldrop, G.L. Movement of the Biotin Carboxylase B-domain as a Result of ATP Binding. *The Journal of Biological Chemistry* **275**, 16183-16190 (2000).
241. Broussard, T., Pakhomova, S., Neau, D.B., Bonnot, R. & Waldrop, G.L. Structural Analysis of Substrate, Reaction Intermediate and Product Binding in Haemophilus influenzae Biotin Carboxylase. *Biochemistry* **54**, 3860-3870 (2015).
242. Cho, Y.S. et al. Crystal structure of the biotin carboxylase domain of human acetyl-CoA carboxylase 2. *Proteins: Structure, Function, and Bioinformatics* **70**, 268-272 (2007).
243. Bernadó, P., Mylonas, E., Petoukhov, M.V., Blackledge, M. & Svergun, D.I. Structural Characterization of Flexible Proteins Using Small-Angle X-ray Scattering. *Journal of the American Chemical Society* **129**, 5656-5664 (2007).
244. Petoukhov, M.V. et al. New developments in the ATSAS program package for small-angle scattering data analysis. *Journal of Applied Crystallography* **45**, 342-350 (2012).
245. Tria, G., Mertens, H.D.T., Kachala, M. & Svergin, D.I. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering *IUCrJ* **2**, 207-217 (2015).
246. Gavalda, S. et al. The Pks13/FadD32 crosstalk for the biosynthesis of mycolic acids in Mycobacterium tuberculosis. *J Biol Chem* **284**, 19255-64 (2009).
247. Staunton, J. & Weissman, K.J. Polyketide biosynthesis: a millennium review *Natural Product Reports* **18**, 380-416 (2001).
248. Maier, T., Leibundgut, M., Boehringer, D. & Ban, N. Structure and function of eukaryotic fatty acid synthases. *Quarterly Reviews of Biophysics* **43**, 373-422 (2010).
249. Patel, M.S., Nemeria, N.S., Furey, W. & Jordan, F. The Pyruvate Dehydrogenase Complexes: Structure-based Function and Regulation. *Journal of Biological Chemistry* **289**, 16615-16623 (2014).
250. Maier, T., Jenni, S. & Ban, N. Architecture of Mammalian Fatty Acid Synthase at 4.5 A Resolution. *Science* **311**, 1258-1262 (2006).
251. Anand, S. et al. SBSPKS: structure based sequence analysis of polyketide synthases. *Nucleic Acids Research* **38**, W487-W496 (2010).
252. Radford, S.E., Laue, E.D., Perham, R.N., Martin, S.R. & Appella, E. Conformational flexibility and folding of synthetic peptides representing an interdomain segment of polypeptide chain in the pyruvate dehydrogenase multienzyme complex of Escherichia coli. *The Journal of Biological Chemistry* **264**, 767-775 (1988).
253. Perham, R.N. Swinging Arms and Swinging Domains in Multifunctional Enzymes: Catalytic Machines for Multistep Reactions. *Annual Reviews of Biochemistry* **69**, 961-1004 (2000).
254. Brignole, E.J., Smith, S. & Asturias, F.J. Conformational flexibility of metazoan fatty acid synthase enables catalysis. *Nat Struct Mol Biol* **16**, 190-197 (2009).

255. Kabsch, W. Automatic indexing of rotation diffraction patterns. *Journal of Applied Crystallography* **21**, 67-72 (1988).
256. Kabsch, W. XDS. *Acta Crystallographica Section D* **66**, 125-132 (2009).
257. Collaborative Comuptational Project, N. The CCP4 suite: programs for protein crystallography. *Acta Crystallographica Section D* **50**, 760-763 (1994).
258. Winn, M.D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D* **67**, 235-242 (2011).
259. McCoy, A.J. et al. Phaser crystallographic software. *Journal of Applied Crystallography* **40**, 658-674 (2007).
260. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D* **60**, 2126-2131 (2004).
261. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallographica Section D* **66**, 486-501 (2010).
262. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D* **66**, 213-221 (2010).
263. Vagin, A. & Teplyakov, A. MOLREP: an Automated Program for Molecular Replacement. *Journal of Applied Crystallography* **30**, 1022-1025 (1997).
264. Chen, V.B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D* **66**, 12-21 (2009).
265. Marabini, R. et al. Xmipp: An Image Processing Package for Electron Microscopy. *Journal of Structural Biology* **116**, 237-240 (1996).
266. Scheres, S.H.W. et al. Modeling Experimental Image Formation for Likelihood-Based Classification of Electron Microscopy Data. *Structure* **15**, 1167-1177 (2007).
267. Schneider, C.A., Rasband, W.S. & Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**, 671-675 (2012).
268. Tang, G. et al. EMAN2: An extensible image processing suite for electron microscopy. *Journal of Structural Biology* **157**, 38-46 (2007).
269. Svergun, D.I., Barberato, C. & Koch, M.H.J. CRYSOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *Journal of Applied Crystallography* **28**, 768-773 (1995).
270. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195-201 (2006).
271. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crytsalline state. *Journal of Molecular Biology* **372**, 774-797 (2007).
272. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792-1797 (2004).
273. Halgren, T. New Method for Fast and Accurate Binding-site Identification and Analysis. *Chemical Biology & Drug Design* **69**, 146-148 (2007).
274. Halgren, T.A. Identifying and Characterizing Binding Sites and Assessing Druggability. *Journal of Chemical Information and Modeling* **49**, 377-389 (2009).
275. Pei, J. & Grishin, N.V. AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics* **17**, 700-712 (2001).
276. Reddy, M.C.M. et al. Structure, Activity, and Inhibition of the Carboxyltransferase -Subunit of Acetyl Coenzyme A Carboxylase

(AccD6) from Mycobacterium tuberculosis. *Antimicrobial Agents and Chemotherapy* **58**, 6122-6132 (2014).
277. Tran, T.H. et al. Structure and function of a single-chain, multi-domain long-chain acyl-CoA carboxylase. *Nature* **518**, 120-123 (2015).
278. Schmid, F.X., Mayr, L.M., Mucke, M. & Schonbrunner, E.R. Prolyl isomerases: role in protein folding. *Adv Protein Chem* **44**, 25-66 (1993).
279. Lang, K., Schmid, F.X. & Fischer, G. Catalysis of protein folding by prolyl isomerase. *Nature* **329**, 268-70 (1987).
280. Siekierka, J.J., Hung, S.H., Poe, M., Lin, C.S. & Sigal, N.H. A cytosolic binding protein for the immunosuppressant FK506 has peptidyl-prolyl isomerase activity but is distinct from cyclophilin. *Nature* **341**, 755-7 (1989).
281. Lu, K.P. & Zhou, X.Z. The prolyl isomerase PIN1: a pivotal new twist in phosphorylation signalling and disease. *Nat Rev Mol Cell Biol* **8**, 904-16 (2007).
282. Harris, G.S., White, T.C., Flory, J.E. & Orme-Johnson, W.H. Genes required for formation of the apoMoFe protein of Klebsiella pneumoniae nitrogenase in Escherichia coli. *J Biol Chem* **265**, 15909-19 (1990).
283. Emelyanov, V.V. & Loukianov, E.V. A 29.5 kDa heat-modifiable major outer membrane protein of Rickettsia prowazekii, putative virulence factor, is a peptidyl-prolyl cis/trans isomerase. *IUBMB Life* **56**, 215-9 (2004).
284. Hodak, H. et al. The peptidyl-prolyl isomerase and chaperone Par27 of Bordetella pertussis as the prototype for a new group of parvulins. *J Mol Biol* **376**, 414-26 (2008).
285. Kakeshita, H. et al. Secretion of biologically-active human interferon-beta by Bacillus subtilis. *Biotechnol Lett* **33**, 1847-52 (2011).
286. Jacobs, M., Andersen, J.B., Kontinen, V. & Sarvas, M. Bacillus subtilis PrsA is required in vivo as an extracytoplasmic chaperone for secretion of active enzymes synthesized either with or without pro-sequences. *Mol.Microbiol.* **8**, 957-966 (1993).
287. Tossavainen, H. et al. NMR solution structure and characterization of substrate binding site of the PPIase domain of PrsA protein from Bacillus subtilis. *FEBS Lett* **580**, 1822-6 (2006).
288. Savitsky, P. et al. High-throughput production of human proteins for crystallization: the SGC experience. *J Struct Biol* **172**, 3-13 (2010).
289. Schmid, F.X. Prolyl isomerase: enzymatic catalysis of slow protein-folding reactions. *Annu Rev Biophys Biomol Struct* **22**, 123-42 (1993).
290. Zoldak, G. et al. A library of fluorescent peptides for exploring the substrate specificities of prolyl isomerases. *Biochemistry* **48**, 10423-36 (2009).
291. Buchner, J. et al. GroE facilitates refolding of citrate synthase by suppressing aggregation. *Biochemistry* **30**, 1586-91 (1991).
292. Myers, J.K., Pace, C.N. & Scholtz, J.M. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci* **4**, 2138-48 (1995).
293. Kather, I., Jakob, R., Dobbek, H. & Schmid, F.X. Changing the determinants of protein stability from covalent to non-covalent interactions by in vitro evolution: a structural and energetic analysis. *J Mol Biol* **381**, 1040-54 (2008).
294. Pervushin, K., Riek, R., Wider, G. & Wuthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift

anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci U S A* **94**, 12366-71 (1997).
295. Guntert, P., Dotsch, V., Wider, G. & Wuthrich, K. Processing of multi-dimensional NMR data with the new software PROSA. *J Biomol NMR* **2**, 619-29 (1992).
296. Keller, R.L.J. The Computer Aided Resonance Assignment Tutorial. *Cantina Verlag, Goldau, 2004* (2004).
297. Zwart, P.H. et al. Automated structure solution with the PHENIX suite. *Methods Mol Biol* **426**, 419-35 (2008).
298. Terwilliger, T.C. et al. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr D Biol Crystallogr* **64**, 61-9 (2008).
299. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* **62**, 1002-11 (2006).
300. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32 (2004).
301. Adams, P.D. et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **58**, 1948-54 (2002).
302. Leskela, S., Wahlstrom, E., Kontinen, V.P. & Sarvas, M. Lipid modification of prelipoproteins is dispensable for growth but essential for efficient protein secretion in Bacillus subtilis: characterization of the Lgt gene. *Mol Microbiol* **31**, 1075-85 (1999).
303. Grasberger, B., Minton, A.P., DeLisi, C. & Metzger, H. Interaction between proteins localized in membranes. *Proc Natl Acad Sci U S A* **83**, 6258-62 (1986).
304. Lee, D., Hilty, C., Wider, G. & Wuthrich, K. Effective rotational correlation times of proteins from NMR relaxation interference. *J Magn Reson* **178**, 72-6 (2006).
305. Fernandez, C., Adeishvili, K. & Wuthrich, K. Transverse relaxation-optimized NMR spectroscopy with the outer membrane protein OmpX in dihexanoyl phosphatidylcholine micelles. *Proc Natl Acad Sci U S A* **98**, 2358-63 (2001).
306. Chou, J.J., Baber, J.L. & Bax, A. Characterization of phospholipid mixed micelles by translational diffusion. *J Biomol NMR* **29**, 299-308 (2004).
307. Jakob, R.P., Zoldak, G., Aumuller, T. & Schmid, F.X. Chaperone domains convert prolyl isomerases into generic catalysts of protein folding. *Proc Natl Acad Sci U S A* **106**, 20282-7 (2009).
308. Heikkinen, O. et al. Solution structure of the parvulin-type PPIase domain of Staphylococcus aureus PrsA--implications for the catalytic mechanism of parvulins. *BMC Struct Biol* **9**, 17 (2009).
309. Scholz, C., Stoller, G., Zarnt, T., Fischer, G. & Schmid, F.X. Cooperation of enzymatic and chaperone functions of trigger factor in the catalysis of protein folding. *EMBO J* **16**, 54-8 (1997).
310. Hottenrott, S., Schumann, T., Pluckthun, A., Fischer, G. & Rahfeld, J.U. The Escherichia coli SlyD is a metal ion-regulated peptidyl-prolyl cis/trans-isomerase. *J Biol Chem* **272**, 15697-15701 (1997).
311. Scholz, C., Rahfeld, J., Fischer, G. & Schmid, F.X. Catalysis of protein folding by parvulin. *J Mol Biol* **273**, 752-62 (1997).

312. Uchida, T., Fujimori, F., Tradler, T., Fischer, G. & Rahfeld, J.U. Identification and characterization of a 14 kDa human protein as a novel parvulin-like peptidyl prolyl cis/trans isomerase. *FEBS Lett* **446**, 278-82 (1999).
313. Schmidpeter, P.A., Jahreis, G., Geitner, A.J. & Schmid, F.X. Prolyl isomerases show low sequence specificity toward the residue following the proline. *Biochemistry* **50**, 4796-803 (2011).
314. Kiefhaber, T., Grunert, H.P., Hahn, U. & Schmid, F.X. Replacement of a cis proline simplifies the mechanism of ribonuclease T1 folding. *Biochemistry* **29**, 6475-80 (1990).
315. Mayr, L.M., Odefey, C., Schutkowski, M. & Schmid, F.X. Kinetic analysis of the unfolding and refolding of ribonuclease T1 by a stopped-flow double-mixing technique. *Biochemistry* **35**, 5550-61 (1996).
316. Mucke, M. & Schmid, F.X. Enzymatic catalysis of prolyl isomerization in an unfolding protein. *Biochemistry* **31**, 7848-54 (1992).
317. Mucke, M. & Schmid, F.X. Folding mechanism of ribonuclease T1 in the absence of the disulfide bonds. *Biochemistry* **33**, 14608-19 (1994).
318. Behrens, S., Maier, R., de Cock, H., Schmid, F.X. & Gross, C.A. The SurA periplasmic PPIase lacking its parvulin domains functions in vivo and has chaperone activity. *EMBO J* **20**, 285-94 (2001).
319. Li, Z. et al. The structure of the Candida albicans Ess1 prolyl isomerase reveals a well-ordered linker that restricts domain mobility. *Biochemistry* **44**, 6180-9 (2005).
320. Landrieu, I., Wieruszeski, J.M., Wintjens, R., Inze, D. & Lippens, G. Solution structure of the single-domain prolyl cis/trans isomerase PIN1At from Arabidopsis thaliana. *J Mol Biol* **320**, 321-32 (2002).
321. Kuhlewein, A. et al. Solution structure of Escherichia coli Par10: The prototypic member of the Parvulin family of peptidyl-prolyl cis/trans isomerases. *Protein Sci* **13**, 2378-87 (2004).
322. Bitto, E. & McKay, D.B. Crystallographic structure of SurA, a molecular chaperone that facilitates folding of outer membrane porins. *Structure* **10**, 1489-98 (2002).
323. Giuseppe, P.O., Von Atzingen, M., Nascimento, A.L., Zanchin, N.I. & Guimaraes, B.G. The crystal structure of the leptospiral hypothetical protein LIC12922 reveals homology with the periplasmic chaperone SurA. *J Struct Biol* **173**, 312-22 (2011).
324. Weininger, U., Jakob, R.P., Kovermann, M., Balbach, J. & Schmid, F.X. The prolyl isomerase domain of PpiD from Escherichia coli shows a parvulin fold but is devoid of catalytic activity. *Protein Sci* **19**, 6-18 (2010).
325. Mueller, J.W. et al. Crystallographic proof for an extended hydrogen-bonding network in small prolyl isomerases. *J Am Chem Soc* **133**, 20096-9 (2011).
326. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-97 (2007).
327. Lill, R., Crooke, E., Guthrie, B. & Wickner, W. The Trigger Factor Cycle Includes Ribosomes, Presecretory Proteins, and the Plasma-Membrane. *Cell* **54**, 1013-1018 (1988).
328. Stoller, G. et al. A Ribosome-Associated Peptidyl-Prolyl Cis/Trans Isomerase Identified as the Trigger Factor. *Embo Journal* **14**, 4939-4948 (1995).
329. Ferbitz, L. et al. Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature* **431**, 590-6 (2004).

330. Goemans, C., Denoncin, K. & Collet, J.F. Folding mechanisms of periplasmic proteins. *Biochim Biophys Acta* **1843**, 1517-28 (2014).
331. Kovermann, M., Schmid, F.X. & Balbach, J. Molecular function of the prolyl cis/trans isomerase and metallochaperone SlyD. *Biol Chem* **394**, 965-75 (2013).
332. Kakeshita, H., Kageyama, Y., Ara, K., Ozaki, K. & Nakamura, K. Propeptide of Bacillus subtilis amylase enhances extracellular production of human interferon-alpha in Bacillus subtilis. *Appl Microbiol Biotechnol* **89**, 1509-17 (2011).
333. Freudl, R. Leaving home ain't easy: protein export systems in Gram-positive bacteria. *Res Microbiol* **164**, 664-74 (2013).
334. Matias, V.R. & Beveridge, T.J. Cryo-electron microscopy reveals native polymeric cell wall structure in Bacillus subtilis 168 and the existence of a periplasmic space. *Mol Microbiol* **56**, 240-51 (2005).
335. Sarvas, M., Harwood, C.R., Bron, S. & van Dijl, J.M. Post-translocational folding of secretory proteins in Gram-positive bacteria. *Biochim Biophys Acta* **1694**, 311-27 (2004).
336. Cahoon, L.A. & Freitag, N.E. Listeria monocytogenes virulence factor secretion: don't leave the cell without a chaperone. *Front Cell Infect Microbiol* **4**, 13 (2014).
337. Burmann, B.M., Wang, C. & Hiller, S. Conformation and dynamics of the periplasmic membrane-protein-chaperone complexes OmpX-Skp and tOmpA-Skp. *Nat Struct Mol Biol* **20**, 1265-72 (2013).
338. Ceh-Pavia, E., Spiller, M.P. & Lu, H. Folding and biogenesis of mitochondrial small Tim proteins. *Int J Mol Sci* **14**, 16685-705 (2013).
339. Eisenberg, D., Schwarz, E., Komaromy, M. & Wall, R. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J Mol Biol* **179**, 125-42 (1984).
340. Forster, B.M. & Marquis, H. Protein transport across the cell wall of monoderm Gram-positive bacteria. *Mol Microbiol* **84**, 405-13 (2012).
341. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* **38**, W529-33 (2010).
342. Jain, E. et al. Infrastructure for the life sciences: design and implementation of the UniProt website. *BMC Bioinformatics* **10**, 136 (2009).
343. Corpet, F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* **16**, 10881-90 (1988).
344. Gouet, P., Robert, X. & Courcelle, E. ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res* **31**, 3320-3 (2003).
345. Wei, J. & Tong, L. Crystal structure of the 500-kDa yeast acetyl-CoA carboxylase holoenzyme dimer. *Nature* **526**, 723-7 (2015).
346. Yu, L.P. et al. A symmetrical tetramer for S. aureus pyruvate carboxylase in complex with coenzyme A. *Structure* **17**, 823-32 (2009).
347. Lietzan, A.D. et al. Interaction between the biotin carboxyl carrier domain and the biotin carboxylase domain in pyruvate carboxylase from Rhizobium etli. *Biochemistry* **50**, 9708-23 (2011).
348. Maier, T., Jenni, S. & Ban, N. Architecture of mammalian fatty acid synthase at 4.5 A resolution. *Science* **311**, 1258-62 (2006).
349. Jenni, S. et al. Structure of fungal fatty acid synthase and implications for iterative substrate shuttling. *Science* **316**, 254-61 (2007).

350. Leibundgut, M., Jenni, S., Frick, C. & Ban, N. Structural basis for substrate delivery by acyl carrier protein in the yeast fatty acid synthase. *Science* **316**, 288-90 (2007).
351. Aevarsson, A., Seger, K., Turley, S., Sokatch, J.R. & Hol, W.G. Crystal structure of 2-oxoisovalerate and dehydrogenase and the architecture of 2-oxo acid dehydrogenase multienzyme complexes. *Nat Struct Biol* **6**, 785-92 (1999).